~~Low CR Candidate~~

AIAA

Reprint

CR

LECTURES ON CONTROLLABILITY AND OBSERVABILITY*

by

R. E. KALMAN**

Stanford University
Stanford, Calif. 94305, U.S.A.
and
Centre d'Automatique
École Nationale Supérieure des Mines de Paris
Paris, FRANCE

Published in Italy

* Lectures delivered at CENTRO INTERNAZIONALE MATEMATICO ESTIVO (C.I.M.E.) Seminar on Controllability and Observability, Fondazione Guglielmo Marconi, Pontecchio Marconi (Bologna, ITALY), from July 1 through July 9, 1968.

** This research was supported in part by Contract AF 49 (638)-1440 and in part by Grant NGR-05-020-073. The writer is indebted to C.I.M.E. for a small travel grant.

CENTRO INTERNAZIONALE MATEMATICO ESTIVO

(C.I.M.E.)

LECTURES ON CONTROLLABILITY AND OBSERVABILITY

R.E. KALMAN

(Stanford-University)

Corso tenuto a Sasso Marconi (Bologna) dal 1 al 9 Luglio 1968

Page intentionally left blank

TABLE OF CONTENTS

R.E. Kalman

INTRODUCTION

The theory of controllability and observability has been developed, one might almost say reluctantly, in response to problems generated by technological science, especially in areas related to control, communication, and computers. It seems that the first conscious steps to formalize these matters as a separate area of (system-theoretic or mathematical) research were undertaken only as late as 1959, by KALMAN [1960b-c]. There have been, however, many scattered results before this time (see Section 12 for some historical comments and references), and one might confidently assert today that some of the main results have been discovered, more or less independently, in every country which has reached an advanced stage of "development" and it is certain that these same results will be rediscovered again in still more places as other countries progress on the road to development.

With the perspective afforded by ten years of happenings in this field, we ought not hesitate to make some guesses of the significance of what has been accomplished. I see two main trends:

(i) The use of the concepts of controllability and observability to study nonclassical questions in optimal control and optimal estimation theory, sometimes as basic hypotheses securing existence, more often as seemingly technical conditions which allow a sharper statement of results or shorter proofs.

(ii) Interaction between the concepts of controllability and observability and the study of structure of dynamical systems, such as: formulation and solution of the problem of realization, canonical forms, decomposition of systems.

The first of these topics is older and has been studied primarily from the point of view of analysis, although the basic lemma (2.7) is purely algebraic. The second group of topics may be viewed as "blowing up" the ideas inherent in the basic lemma (2.7), resulting in a more and more strictly algebraic point of view.

There is active research in both areas.

In the first, attention has shifted from the case of systems governed by finite-dimensional linear differential equations with constant coefficients (where success was quick and total) to systems governed by infinite-dimensional linear differential equations (delay differential equations, classical types of partial differential equations, etc.), to finite-dimensional linear differential equations with time-dependent coefficients, and finally to all sorts and subsorts of nonlinear differential equations. The first two topics are surveyed concurrently by WEISS [1969] while MARKUS [1965] looks at the nonlinear situation.

My own current interest lies in the second stream, and these lectures will deal primarily with it, after a rather hurried overview of the general problem and of the "classical" results.

Let us take a quick look at the most important of these "classical" results. For convenience I shall describe them in system-theoretic (rather than conventional pure mathematical) language. The mathematically trained reader should have no difficulty in converting them into his preferred framework, by digging a little into the references.

In area (i), the most important results are probably those which give more or less explicit and computable results for controllability and observability of certain specific classes of systems. Beyond these, there seem to be two main theorems:

THEOREM A. A real, continuous-time, n-dimensional, constant, linear dynamical system Σ has the property "every set of n eigenvalues may be produced by suitable state feedback" if and only if Σ is completely controllable.

The central special case is treated in great detail by KALMAN, FALB, and ARBIB [1969, Chapter 2, Theorem 5.10]; for a proof of the general case with background comments, refer to WONHAM [1967]. As a particular case, we have that every system satisfying the hypotheses of the theorem can be "stabilized" (made to have eigenvalues with negative real parts) via a suitable choice of feedback. This result is the "existence theorem" for algorithms used to construct control systems for the past three decades, and yet a conscious formulation of the problem and its mathematical solution go back to about 1963! (See Theorem D below.) The analogous problem for nonconstant linear systems (governed by linear differential equations with variable coefficients) is still not solved.

THEOREM B. ("Duality Principle") *Every problem of controllability in a real, (continuous-time, or discrete-time), finite-dimensional, constant, linear dynamical system is equivalent to a controllability problem in a dual system.*

This fact was first observed by KALMAN [1960a] in the solution of the optimal stochastic filtering problem for discrete-time systems, and was soon applied to several problems in system theory by KALMAN [1960b-c]. See also many related comments by KALMAN, FALB, and ARBIB [Chapters 2 and 6, 1969]. As a theorem, this principle is not yet known to be valid outside the linear area, but as an intuitive prescription it has been rather useful in guiding system-theoretic research. The problems involved here are those of fomulation rather than proof. The basic difficulties seem to point toward algebra and in particular category theory. System-theoretic duality, like the categoric one, is concerned with "reversing arrows". See Section 10 for a modern discussion of these points and a precise version of Theorem B.

Partly as a result of the questions raised by Theorem B and partly because of the algebraic techniques needed to prove Theorem A and related lemmas, attention in the early 1960's shifted toward certain problems of a structural nature which were, somewhat surprisingly at first, found to be related to controllability and observability. The main theorems again seem to be two:

THEOREM C. (Canonical Decomposition) *Every real (continuous-time or discrete-time), finite-dimensional, constant, linear dynamical* system may be canonically decomposed into four parts, of which only one part, that which is completely controllable and completely observable, is involved in the input/output behavior of the system.

The proof given by KALMAN [1962] applies to nonconstant systems only under the severe restriction that the dimensions of the subspace of all controllable and all unobservable states is constant on the whole real line. The result represented by Theorem C is far from definitive, however, since finite-dimensional linear, nonconstant systems admit at least four different canonical decompositions: it is possible and fruitful to dualize the notions of controllability and observability, thereby arriving at four properties, presently called

reachability and controllability

as well as

constructibility* and observability.

(See Section 2 definitions.) Any combination of a property from the first list with a property from the second list gives a canonical decomposition result analogous to Theorem C. The complexity of the situation was first revealed by WEISS and KALMAN [1965]; this paper contributed to a revival of interest (with hopes of success) in the special problems of nonconstant linear systems. Recent

---

*WEISS [1969] uses "determinability" instead of constructibility. The new terminology used in these lectures is not yet entirely standard.

progress is surveyed by WEISS [1969]. Intimately related to the canonical structure theorem, and in fact necessary to fully clarify the phrase "involved in the input/output behavior of the system", is the last basic result:

THEOREM D. (Uniqueness of Minimal Realization) *Given the impulse-response matrix* W *of a real, continuous-time, finite-dimensional, linear dynamical system, there exists a real, continuous-time, finite-dimensional, linear dynamical system* Σ_W *which*

(a) *realizes* W: *that is, the impulse-response matrix of* Σ_W *is equal to* W;

(b) *has minimal dimension in the class of linear systems satisfying* (a);

(c) *is completely controllable and completely observable*;

(d) *is uniquely determined (modulo the choice of a basis at each* t *for its state space) by requirement* (a) *together with* (b) *or, independently, by* (a) *together with* (c).

In short, for any W *as described above, there is an "essentially unique"* Σ_W *of the same "type" which satisfies* (a) *through* (c).

COROLLARY 1. *If* W *comes from a constant system, there is a constant* Σ_W *which satisfies* (a) *through* (c), *and is uniquely determined by* (a) + (b) *or* (a) + (c) (*modulo a fixed choice of basis for its state space*).

COROLLARY 2. All claims of Corollary 1 continue to hold if "impulse-response matrix of a constant, finite-dimensional system" is replaced by "transfer function matrix of a constant, finite-dimensional system".

The first general discussion of the situation with an equivalent statement of Theorem D is due to KALMAN [1963b, Theorems 7 and 8]. (This paper does not include complete proofs, or even an explicit statement of Corollaries 1 and 2, although they are implied by the general algorithm given in Section 7. An edited version of the original unpublished proof of Theorem D is given in KALMAN, FALB, and ARBIB [1969, Chapter 10, Appendix C].)

These results are of great importance in engineering system theory since they relate methods based on the Laplace transform (using the transfer function of the system) and the time-domain methods based on input/output data (the matrix W) to the state-variable (dynamical system) methods developed in 1955-1960. In fact, by Corollary 1 it follows that the two methods must yield identical results; for instance, starting with a constant impulse-response matrix W, property (c) implies that the existence of a stable control lay is always assured by virtue of Theorem A. Thus it is only after the development represented by Theorems A-D that a rigorous justification is obtained for the intuitive design methods used in control engineering.

As with Theorem C, certain formulational difficulties arise in connection with a precise definition of a "nonconstant linear dynamical system". Thus, it seems preferable at present to replace in Theorem D "impulse-response matrix W" by "weighting pattern W" (or "abstract input/output map W") and "complete controllability" by "complete reachability". The definitive form of the 1963 theorem evolved through the works of WEISS and KALMAN [1965], YOULA [1966], and KALMAN; a precise formulation and modernized proof of Theorem D in the weighting pattern case was given recently by KALMAN, FALB, and ARBIB [1969, Chapter 10, Section 13.] A completely general discussion of what is meant by a "minimal realization" of a non-constant impulse-response matrix involves many technical complications due to the fact that such a minimal realization does not exist in the class of linear differential equations with "nice" coefficient functions. For the current status of this problem, consult especially DESOER and VARAIYA [1967], SILVERMAN and MEADOWS [1969], KALMAN, FALB, and ARBIB [1969, Chapter 10, Section 13] and WEISS [1969].

From the standpoint of the present lectures, by far the most interesting consequence of Theorem D is its influence, via efforts to arrive at a definitive proof of Corollary 1, on the development of the algebraic stream of system theory. The first proof of this important result (in the special case of distinct eigenvalues) is that of GILBERT [1963]. Immediately afterwards, a general proof was given by KALMAN [1963b, Section 7]. This proof, strictly computational and linear algebraic in nature, yields no theoretical insight although it is useful as the basis of a computer algorithm.

Using the classical theory of invariant factors, KALMAN [1965a] succeeded in showing that the solution of the minimal realization problem can be effectively reduced to the classical invariant-factor algorithm. This result is of great theoretical interest since it strongly suggests the now standard module theoretic approach, but it does not lead to a simple proof of Corollary 1 and is not a practical method of computation.

The best known proof of Corollary 1 was obtained in 1965 by B. L. Ho, with the aid of a remarkable algorithm, which is equally important from a theoretical and computational viewpoint. The early formulation of the algorithm was described by HO and KALMAN [1966], with later refinements discussed in HO and KALMAN [1969], KALMAN, FALB, and ARBIB [1969, Chapter 10, Section 11] and KALMAN [1969c]. Almost simultaneously with the work of B. L. Ho, the basic results were discovered independently also by YOULA and TISSI [1966] and by SILVERMAN [1966]. The subject goes back to the 19th century and centers around the theory of Hankel matrices; however, many of the results just referenced seem to be fundamentally new. This field is currently in a very active stage of development. We shall discuss the essential ideas involved in Sections 8-9. Many other topics, especially Silverman's generalization of the algorithm to nonconstant systems unfortunately cannot be covered due to lack of time.

Acknowledgment

It is a pleasure to thank C. I. M. E. and its organizers, especially Professors E. Bompiani, E. Sarti, and E. Belardinelli, for arranging a special conference on these topics. The sunny skies and hospitality of Italy, along with Bolognese food played a subsidiary but vital part in the success of this important gathering of scientists.

R.E.Kalman

1. CLASSICAL AND MODERN DYNAMICAL SYSTEMS

In mathematics the term dynamical system (synonyms: topological dynamics, flows, abstract dynamics, etc.) usually connotes the action of a one-parameter group T (the reals) on a set X, where X is at least a topological space (more often, a differentiable manifold) and the action is at least continuous. This setup is physically motivated, but in a very old-fashioned sense. A "dynamical system" as just defined is an idealization, generalization, and abstraction of Newton's world view of the Solar System as described via a finite set of nonlinear ordinary differential equations. These equations represent the positions and momenta of the planets regarded as point masses and are completely determined by the laws of gravitation, i.e., they do not contain any terms to account for "external" forces that may act on the system.

Interesting as this notation of a dynamical system may be (and is!) in pure mathematics, it is much too limited for the study of those dynamical systems which are of contemporary interest. There are at least three different ways in which the classical concept must be generalized:

(i) The time set of the system is not necessarily restricted to the reals;

(ii) A state $x \in X$ of the system is not merely acted upon by the "passage of time" but also by inputs which are or could be manipulated to bring about a desired type of behavior;

(iii) The states of the system cannot, in general, be observed. Rather, the physical behavior of the system is manifested through its *outputs* which are many-to-one functions of the state.

The generalization of the time set is of minor interest to us here. The notions of *input* and *output*, however, are exceedingly fundamental; in fact, controllability is related to the input and observability to the output. With respect to dynamical systems in the classical sense, neither controllability nor observability are meaningful concepts.

A much more detailed discussion of dynamical systems in the modern sense, together with rather detailed precise definitions, will be found in KALMAN, FALB, and ARBIB [1969, Chapter 1].

From here on, we will use the term "dynamical system" exclusively in the modern sense (we have already done so in the Introduction).

The following symbols will have a fixed meaning throughout the paper:

$$(1.1)\quad \begin{cases} T = \text{time set}, \\ U = \text{set of input values}, \\ X = \text{state set}, \\ Y = \text{set of output values}, \\ \Omega = \text{input functions}, \\ \varphi = \text{transition map}, \\ \eta = \text{readout map}. \end{cases}$$

The following assumptions will always apply (otherwise the sets above are arbitrary):

$$(1.2)\quad \left\{\begin{array}{lll} T & = & \text{an ordered subset of the reals } \underline{\underline{R}}, \\ \Omega & = & \text{class of functions } T \to U \text{ such that} \\ & & \text{(i) each function } \omega \text{ is undefined outside some} \\ & & \qquad \text{finite interval } J_\omega \subset T \text{ dependent on } \omega; \\ & & \text{(ii) if } J_\omega \cap J_{\omega'} = \emptyset, \text{ there is a function} \\ & & \qquad \omega \in \Omega \text{ which agrees with } \omega \text{ on } J_\omega \text{ and} \\ & & \qquad \text{with } \omega' \text{ on } J_{\omega'}. \end{array}\right.$$

For most purposes later, T will be equal to $\underline{\underline{Z}}$ = (ordered) abelian group of integers; U, X, Y, Ω will be linear spaces; "undefined" can be replaced by "equal to 0"; and "functions undefined outside a finite interval" will mean the same as "finite sequences".

The most general notion of a dynamical system for our present needs is given by the following

(1.3) DEFINITION. *A dynamical system* Σ *is a composite object consisting of the maps* φ, η *defined on the sets* T, U, Ω, X, Y *(as above)*:

$$\varphi\colon\ T \times T \times X \times \Omega \to X,$$
$$:\ (t;\ \tau,\ x,\ \omega) \mapsto \varphi(t;\ \tau,\ x,\ \omega)$$

undefined whenever $t \geqq \tau$;

$$\eta\colon\ T \times X \to Y\colon\ (t,\ x) \mapsto \eta(t,\ x).$$

The transition map φ *satisfies the following assumptions*:

(1.4) $\varphi(t;\ t,\ x,\ \omega) = x;$

(1.5) $\varphi(t; \tau, x, \omega) = \varphi(t; s, \varphi(s; \tau, x, \omega), \omega)$;

(1.6) <u>if</u> $\omega = \omega'$ <u>on</u> $[\tau, t]$, <u>then for all</u> $s \in [\tau, t]$

$\varphi(s; \tau, x, \omega) = \varphi(s; \tau, x, \omega')$.

The definition of a dynamical system on this level of generality should be regarded only as a scaffolding for the terminology; interesting mathematics begins only after further hypotheses are made. For instance, it is usually necessary to endow the sets T, U, Ω, X, and Y with a topology and then require that φ and η be continuous.

(1.7) EXAMPLE. The classical setup in topological dynamics may be deduced from our Definition (1.3) in the following way. Let $T = \underline{\underline{R}}$ = reals, regarded as an abelian group under the usual addition and having the usual topology; let Ω consist only of the nowhere-defined function; let X be topological space; disregard Y and η entirely; define φ for <u>all</u> $t, \tau \in T$ and write it as

$$\varphi(t; \tau, x, \omega) = x\cdot(t - \tau),$$

that is, a function of x and $t - \tau$ alone. Check (1.4-5); in the new notation they become

$$x\cdot 0 = x \quad\text{and}\quad x\cdot(s + t) = (x\cdot s)\cdot t.$$

Finally, require that the map $(x, t) \mapsto x\cdot t$ be continuous.

(1.8) INTERPRETATION. The essential idea of Definition (1.3) is that it axiomatizes the notion of state. A dynamical system is informally a rule for state transitions (the function φ), together with suitable means of expressing the effect of the input on the state and the effect of the state on the output (the function η). The map φ is verbalized as follows: "an input ω, applied to the system Σ in state x at time τ produces the state $\varphi(t; \tau, x, \omega)$ at time t." The peculiar definition of an input function ω is used here mainly for technical convenience; by (1.6) only equivalence classes of inputs agreeing over $[\tau, t]$ enter into the determination of $\varphi(t; \tau, x, \omega)$. "$\omega$ not defined" at t means no input acts on Σ at time t.

The pair $(\tau, x) \in T \times X$ will be called an event of a dynamical system Σ.

In the sequel, we shall be concerned primarily with systems which are finite-dimensional, linear, and continuous-time or discrete-time. Often these systems will be also real and constant (= stationary or time-invariant). We leave the precise definition of these terms in the context of Definition (1.3) to the reader (consult KALMAN, FALB, or ARBIB [1969, Chapter 1] as needed) and proceed to make some ad hoc definitions without detailed explanation.

The following conventions will remain in force throughout the lectures whenever the linear case is discussed:

(1.9) Continuous-time. $T = \underline{\underline{R}}$, $U = \underline{\underline{R}}^m$, $X = \underline{\underline{R}}^n$, $Y = \underline{\underline{R}}^p$, Ω = all continuous functions $\underline{\underline{R}} \to \underline{\underline{R}}^m$ which vanish outside a finite interval.

(1.10) Discrete-time. $T = \underline{\underline{Z}}$, K = fixed field (arbitrary),

$U = K^m$, $X = K^n$, $Y = K^p$, Ω = all functions
$\mathbb{Z} \to K^m$ which are zero for all but a finite number of
their arguments.

Now we have, finally,

(1.11) DEFINITION. *A real, continuous-time, n-dimensional, linear dynamical system* Σ *is a triple of continuous matrix functions of time* $(F(\cdot), G(\cdot), H(\cdot))$ *where*

$$F(\cdot): \mathbb{R} \to \{n \times n \text{ matrices over } \mathbb{R}\}$$

$$G(\cdot): \mathbb{R} \to \{n \times m \text{ matrices over } \mathbb{R}\},$$

$$H(\cdot): \mathbb{R} \to \{p \times n \text{ matrices over } \mathbb{R}\}.$$

These maps determine the equations of motion of Σ *in the following manner:*

$$(1.12) \quad \begin{cases} dx/dt = F(t)x + G(t)\omega(t), \\ y(t) = H(t)x(t), \end{cases}$$

where $t \in \mathbb{R}$, $x \in \mathbb{R}^n$, $\omega(t) \in \mathbb{R}^m$, *and* $y(t) \in \mathbb{R}^p$.

To check that (1.12) indeed makes Σ into a well-defined dynamical system in the sense of Definition (1.3), it is necessary to recall the basic facts about finite systems of ordinary linear differential equations with continuous coefficients. Define the map

$$\Phi_F(t, \tau): \mathbb{R} \times \mathbb{R} \to \{n \times n \text{ matrices over } \mathbb{R}\}$$

to be the family of $n \times n$ matrix solutions of the linear differential equation

$$dx/dt = F(t)x, \quad x \in \underline{\underline{R}}$$

subject to the initial condition

$$\Phi_F(\tau, \tau) = I = \text{unit matrix}, \quad \tau \in \underline{\underline{R}}.$$

Then Φ_F is of class C^1 in both arguments. It is called the *transition matrix* of (the system Σ whose "infinitesimal" transition matrix is) $F(\cdot)$. From this standard result we get easily also the fact that the transition map of Σ is explicitly given by

$$(1.13) \qquad \varphi(t; \tau, x, \omega) = \Phi_F(t, \tau)x + \int_\tau^t \Phi_F(t, s)G(s)G'(s)\Phi'_F(t, s)ds$$

while the *readout map* is given by

$$(1.14) \qquad \eta(t, x) = H(t)x.$$

It is instructive to verify that φ indeed depends only on the equivalence class of ω's which agree on $[\tau, t]$.

In view of the classical terminology "linear differential equations with constant coefficients", we introduce the nonstandard

(1.15) DEFINITION. *A real, continuous-time, finite-dimensional linear dynamical system* $\Sigma = (F(\cdot), G(\cdot), H(\cdot))$ *is called constant iff all three matrix functions are constant.*

In strict analogy with (1.15), we say:

(1.16) DEFINITION. *A discrete-time, finite-dimensional, linear, constant dynamical system* Σ *over* K *is a triple* (F, G, H) *of*

$n \times n$, $n \times m$, $p \times n$ *matrices over the field* K. *These maps determine the equations of motion of* Σ *in the following manner:*

$$(1.17)\quad \begin{cases} x(t+1) = Fx(t) + G\omega(t), \\ y(t) = Hx(t), \end{cases}$$

where $t \in \mathbf{Z}$, $x \in K^n$, $\omega(t) \in K^m$, *and* $y(t) \in K^p$.

In the sequel, we shall use the notations $(F, G, -)$ or $F, -, H)$ to denote systems possessing certain properties which are true for any H or G.

Finally, we adopt the following convention, which is already implicit in the preceding discussion:

(1.18) DEFINITION. *The dimension* n *of a dynamical system* Σ *is equal to the dimension of* X_Σ *as a vector space.*

2. STANDARDIZATION OF DEFINITIONS AND "CLASSICAL" RESULTS

In this section, we shall be mainly interested in finite-dimensional linear dynamical systems, although the first two definitions will be quite general.

Let Σ be an arbitrary dynamical system as defined in Section 1. We assume the following slightly special property: There exists a state x^* and an input ω^* such that

$$\varphi(t;\ \tau,\ x^*,\ \omega^*) \ = \ x^* \quad \text{for all} \quad t,\ \tau \in T \quad \text{and} \quad t \geqq \tau.$$

For simplicity, we write x^* and ω^* as 0. (When X and Ω have additive structure, 0 will have the usual meaning.) The next two definitions refer to dynamical systems with this extra property.

(2.1) DEFINITION. *An event* $(\tau,\ x)$ *is controllable iff*[§] *there exists a* $t \in T$ *and an* $\omega \in \Omega$ (*both* t *and* ω *may depend on* $(\tau,\ x)$) *such that*

$$\varphi(t;\ \tau,\ x,\ \omega) \ = \ 0$$

In words: an event is controllable iff it can be transferre to 0 in finite time by an appropriate choice of the input function ω. Think of the path from $(\tau,\ x)$ to $(t,\ 0)$ as the graph of a function defined over $[\tau,\ t]$.

[§] The technical word iff means if and only if.

Consider now a reflection of this graph about τ. This suggests a new definition which is a kind of "adjoint" of the definition of controllability:

(2.2) DEFINITION. *An event* (τ, x) *is reachable iff there is an* $s \in T$ *and an* $\omega \in \Omega$ (*both* s *and* ω *may depend on* (τ, x)) *such that*

$$x = \varphi(\tau; s, 0, \omega).$$

We emphasize: controllability and reachability are entirely different concepts. A striking example of this fact is encountered below in Proposition (4.26).

We shall now review briefly some well-known criteria for and relations between reachability and controllability in linear systems.

(2.3) PROPOSITION. *In a real, continuous-time, finite-dimensional, linear dynamical system* $\Sigma = (F(\cdot), G(\cdot), - \ )$, *an event* (τ, x) *is*

(a) *reachable if and only if* $x \in \text{range}\ \hat{W}(s, \tau)$ *for some* $s \in \underline{\underline{R}}$, $s < \tau$, *where*

$$\hat{W}(s, \tau) = \int_s^\tau \Phi_F(\tau, \sigma)G(\sigma)G'(\sigma)\Phi_F'(\tau, \sigma)d\sigma$$

(b) *controllable if an only if* $x \in \text{range}\ W(\tau, t)$ *for some* $t \in \underline{\underline{R}}$, $t > \tau$, *where*

$$W(\tau, t) = \int_\tau^t \Phi_F(\tau, s)G(s)G'(s)\Phi_F'(\tau, s)ds.$$

The original proof of (b) is in KALMAN [1960b]; both cases are treated in detail in KALMAN, FALB, and ARBIB [1969, Chapter 2, Section 2]. Note that if $G(\cdot)$ is identically zero on $(-\infty, \tau)$ we cannot have reachability, and if $G(\cdot)$ is identically zero on $(\tau, +\infty)$ we cannot have controllability.

For a constant system, the integrals above depend only on the difference of the limits; hence, in particular

$$W(\tau, t) = \hat{W}(2\tau - t, \tau).$$

So we have

(2.4) PROPOSITION. *In a real, continuous-time, finite-dimensional, linear, constant dynamical system an event* (τ, x) *is reachable for all* τ *if and only if it is reachable for one* τ; *an event is reachable if and only if it is controllable.*

From (2.3) one can obtain in a straightforward fashion also the following much stronger result:

(2.5) THEOREM. *In a real, continuous-time, n-dimensional, linear, constant dynamical system* $\Sigma = (F, G, -)$ *a state* x *is reachable (or, equivalently, controllable) at any* $\tau \in \mathbb{R}$ *if and only if*

$$x \in \operatorname{span}(G, FG, \dots) \subset \mathbb{R}^n;$$

if this condition is satisfied, we can choose $s = \tau - \delta$, $t = \tau + \delta$, *with* $\delta > 0$ *arbitrary*. (The span of a sequence of matrices is to be interpreted as the vector space generated by the columns of these matrices.)

A proof of (2.5) may be found in KALMAN, HO, and NARENDRA [1963] and in KALMAN, FALB, and ARBIB [1969, Chapter 2, Section 3]. A trivial but noteworthy consequence is the fact that the definition of reachable states of Σ is "coördinate-free":

(2.6) COROLLARY. *The set of reachable (or controllable) states of* Σ *in Theorem* (2.5) *is a subspace of the real vector space* X_Σ, *the state space of* Σ.

Very often the attention to individual states is unnecessary and therefore many authors prefer to use the terminology "Σ is completely reachable at τ" for "every event (τ, x), τ = fixed, $x \in X_\Sigma$, is reachable", or "Σ completely reachable" for "every event in Σ is reachable", etc. Thus (2.5), together with the Cayley-Hamilton theorem, implies the

(2.7) BASIC LEMMA. *A real, continuous-time, n-dimensional, linear, constant dynamical system* $\Sigma = (F, G, -)$ *is completely reachable if an only if*

(2.8) $$\operatorname{rank}\,(G, FG, \ldots, F^{n-1}G) = n.$$

Condition (2.8) is very well-known; it or equivalent forms of it have been discovered, explicitly used, or implicitly assumed by many authors. A trivially equivalent form of (2.7) is given by

(2.9) COROLLARY 1. *A constant system* $\Sigma = (F, G, -)$ *is completely reachable if and only if the smallest F-invariant subspace of* X_Σ *containing (all column vectors of)* G *is* X_Σ *itself.*

A useful variant of the last fact is given by

(2.10) COROLLARY 2. (W. Hahn) *A constant system* $\Sigma = (F, G, -)$ *is completely reachable if and only if there is no nonzero eigenvector of* F *which is orthogonal to (every column vector of)* G.

Finally, let us note that, far from being a technical condition, (2.5) has a direct system-theoretic interpretation, as follows:

(2.11) PROPOSITION. *The state space* X_Σ *of a real, continuous-time, n-dimensional, linear, constant dynamical system* $\Sigma = (F, G, -)$ *may be written as a direct sum*

$$X_\Sigma = X_1 \oplus X_2,$$

which induces a decomposition of the equations of motion as (obvious notations)

$$(2.12) \quad \begin{cases} dx_1/dt = F_{11}x_1 + F_{12}x_2 + G_1u(t), \\ dx_2/dt = F_{22}x_2. \end{cases}$$

The subsystem $\Sigma_1 = (F_{11}, G_1, -)$ *is completely reachable. Hence a state* $x = (x_1, x_2) \in X_\Sigma$ *is reachable if and only if* $x_2 = 0$.

PROOF. We define X_1 to be the set of reachable states of Σ; by (2.5) this is an F-invariant subspace of X_Σ. Hence, by finite-dimensionality, X_1 is a direct summand in X_Σ. By construction, every state in X_1 is reachable, and (every column vector of)

G belongs to X_1. The F-invariance of X_1 implies that $F_{11} = 0$, which implies the asserted form of the equations of motion. □

(2.13) REMARK. Note that X_2 is not intrinsically defined (it depends on an arbitrary choice in completing the direct sum). Hence to say that "$(0, x_2)$ is an unreachable (or uncontrollable) state if $x_2 \neq 0$" is an abuse of language. More precisely: *the set of all reachable (or controllable) states has the structure of a vector space, but the set of all unreachable (or uncontrollable) states does not have such structure*. This fact is important to bear in mind for the algebraic development which follows after this section and also in the definition of observability and constructibility below. In general, the direct sum cannot be chosen in such a way that $F_{12} = 0$.

While condition (2.8) has been frequently used as a *technical* requirement in the solution of various optimal control problems in the late 1950's, it was only in 1959-60 that the relation between (2.8) and system theoretic questions was clarified by KALMAN [1960b-c] via Definition (2.2) and Propositions (2.5) and (2.11). (See Section 11 for further details.) In other words, without the preceding discussion the use of (2.8) may appear to be artificial, but in fact it is not, *at least in problems in which control enters*, because, by (2.12) control problems stated for X_Σ are nontrivial only with respect to the intrinsic subspace X_1.

The hypothesis "constant" is by no means essential for Proposition (2.11), but we must forego further comments here.

For later purposes, we state some facts here for discrete-time, constant linear systems analogous to those already developed for their continuous-time counterparts. The proofs are straightforward and therefore omitted (or given later, for illustrative purposes).

(2.14) PROPOSITION. *A state* x *of a real, discrete-time, n-dimensional, linear, constant dynamical system* $\Sigma = (F, G, -)$ *is reachable if and only if*

$$x \in \operatorname{span}\,(G, FG, \ldots, F^{n-1}G). \tag{2.15}$$

Thus such a system is completely reachable if and only if (2.8) *holds.*

(2.16) PROPOSITION. *A state* x *of the system* Σ *described in Proposition* (2.14) *is controllable if and only if*

$$x \in \operatorname{span}\,(F^{-1}G, \ldots, F^{-n}G), \tag{2.17}$$

where

$$F^{-k}G = \{x: \; F^{k}x = g_i, \; g_i = \text{column vector of } G\}.$$

(2.18) PROPOSITION. *In a real, discrete-time, finite-dimensional, linear, constant dynamical system* $\Sigma = (F, G, -)$ *a reachable state is always controllable and the converse is always true whenever* $\det F \neq 0$.

Note also that Propositions (2.11) and its proof continue to be correct, without any modification, when "continuous-time" is replaced by "discrete-time".

Now we turn to a discussion of observability.

The original definition of observability by KALMAN [1960b, Definition (5.23)] was concocted in such a way as to take advantage of vector-space duality. The conceptual problems surrounding duality are easy to handle in the linear case but are still by no means fully understood in the nonlinear case (see Section 10). In order to get at the main facts quickly, we shall consider here only the linear case and even then we shall use the underlying idea of vector-space duality in a rather ad-hoc fashion. The reader wishing to do so can easily turn our remarks into a strictly dual treatment of facts (2.1)-(2.12) with the aid of the setup introduced in Section 10.

(2.19) DEFINITION. *An event* (τ, x) *in a real, continuous-time, finite-dimensional, linear dynamical system* $\Sigma = (F(\cdot), -, H(\cdot))$ *is unobservable iff*

$$H(s)\Phi_F(s, \tau)x = 0 \quad \underline{\text{for all}} \quad s \in [\tau, \infty).$$

(2.20) DEFINITION. *With respect to the same system, an event* (τ, x) *is unconstructible* iff*

----

*In the older literature, starting with KALMAN [1960b, Definition (5.23)], it is *this* concept which is called "observability". By hindsight, the present choice of words seems to be more natural to the writer.

$$H(\sigma)\Phi_F(\sigma, \tau)x = 0 \quad \underline{\text{for all}} \quad \sigma \in (-\infty, \tau].$$

The motivation for the first definition is obvious: the "occurrence" of an unobservable event cannot be detected by looking at the output of the system after time τ. (The definition subsumes $\omega = 0$, but this is no loss of generality because of linearity.) The motivation for the second definition is less obvious but is in fact strongly suggested by statistical filtering theory (see Section 10). In any case, Definition (2.21) complements Definition (2.20) in exactly the same way as Definition (2.1) complements Definition (2.2).

From these definitions, it is very easy to deduce the following criteria:

(2.21) PROPOSITION. *In a real, continuous-time, finite-dimensional, linear dynamical system* $\Sigma = (F(\cdot), -, H(\cdot))$ *an event* (τ, x) *is*

(a) *unobservable if and only if* $x \in \text{kernel } \hat{M}(\tau, t)$ *for all* $t \in \underline{R}$, $t > \tau$, *where*

$$\hat{M}(\tau, t) = \int_\tau^t \Phi_F'(s, \tau)H'(s)H(s)\Phi_F(s, \tau)ds;$$

(b) *unconstructible if and only if* $x \in \text{kernel } M(s, \tau)$ *for all* $s \in \underline{R}$, $s < \tau$, *where*

$$M(s, \tau) = \int_s^\tau \Phi_F'(\sigma, \tau)H'(\sigma)H(\sigma)\Phi_F(\sigma, \tau)d\sigma.$$

PROOF. Part (a) follows immediately from the observation: $x \in \text{kernel } M(\tau, t) \Leftrightarrow H(s)\Phi_F(s, \tau)x = 0$ for all $s \in [\tau, t]$. Part (b) follows by an analogous argument. □

(2.22) REMARK. Let us compare this result with Proposition (2.3), and let us indulge (only temporarily) in abuses of language of the following sort:*

$$(\tau, x) = \underline{\text{unreachable}} \Leftrightarrow x \in \text{kernel } \hat{W}(\tau, t)$$
$$\underline{\text{for all}} \quad t > \tau$$

and

$$(\tau, x) = \underline{\text{observable}} \Leftrightarrow x \in \text{range } \hat{M}(\tau, t)$$
$$\underline{\text{for some}} \quad t > \tau.$$

From these relations we can easily deduce the so-called "duality rules"; that is, problems involving observability (or constructibility) are converted into problems involving reachability (or controllability) in a suitably defined dual system. See KALMAN, FALB, and ARBIB [1969, Chapter 2, Proposition (6.12)] and the broader discussion in Section 10.

We *will* say, by slight abuse of language, that a system is *completely observable* whenever 0 is the only unobservable state. Thus the Basic Lemma (2.7) "dualizes" to the

(2.23) PROPOSITION. *A real, continuous-time or discrete-time, n-dimensional, linear, constant dynamical system* $\Sigma = (F, -, H)$

*All this would be strictly correct if we agreed to replace "direct sum" in Proposition (2.11) and its counterpart (2.25) by "orthogonal direct sum"; but this would be an arbitrary convention which, while convenient, has no natural system-theoretic justification. Reread Remark (2.13).

is completely observable if and only if

(2.24) $\operatorname{rank}\,(H', F'H', \ldots, (F')^{n-1}H') = n.$

By duality, complete constructibility in a continuous-time system is equivalent to observability; in a discrete-time system this is not true in general but it is true when $\det F \neq 0$.

It is easy to see also that (2.11) "dualizes" to:

(2.25) PROPOSITION. _The state space_ X_Σ _of a real, continuous-time or discrete-time , n-dimensional, linear, constant dynamical system_ $\Sigma = (F, -, H)$ _may be written as a direct sum_

$$X_\Sigma = X_1 \oplus X_2$$

and the equations of Σ _are decomposed correspondingly as_

$$\begin{aligned} dx_1/dt &= F_{11}x_1, \\ dx_2/dt &= F_{21}x_1 + F_{22}x_2, \\ y(t) &= H_2x_2(t). \end{aligned}$$

PROOF. Proceed dually to the proof of Proposition (2.11), beginning with the definition of X_1 as the set of all _un_observable states of Σ. □

Combining Propositions (2.11) and (2.25) gives Theorem C as in KALMAN [1962].

This completes our survey of the "classical" results related to reachability, controllability, observability, and constructibility.

The remaining lectures will be concerned exclusively with discrete-time systems. The main motivation for the succeeding developments will be the algebraic criteria (2.8) and (2.24) as well as a deeper examination of Theorems C and D of the Introduction.

3. DEFINITION OF STATES VIA NERODE EQUIVALENCE CLASSES

A classical dynamical system is essentially the action of the time set T (= reals) on the states X. In other words, the states are acted on by an abelian group, namely ($\underline{\underline{R}}$ + usual definition of addition). This is a trivial fact, but it has deep consequences. A (modern) dynamical system is the action of the inputs Ω on X; in exact analogy with the classical case, to the abelian structure on T there corresponds an (associative but noncommutative) semigroup structure on Ω. The idea that Ω always admits such a structure was apparently overlooked until the late 1950's when it became fashionable in automata theory (school of SCHUTZENBERGER). This seems to be the "right" way of translating the intuitive notion of dynamics into mathematics, and it will be fundamental in our succeeding investigations.

It is convenient to assume from now on, until the end of these lectures, that

(3.1) $\quad T$ = time set = $\underline{\underline{Z}}$ = additive (ordered) group of integers.

Since we shall be only interested in constant systems from here on, we shall adopt the following normalization convention:*

---

*In the discrete-time nonconstant case, we would have to deal with $\underline{\underline{Z}}$ copies of Ω, each normalized with respect to a different particular value of $\tau \in \underline{\underline{Z}}$.

(3.2) $\quad$ No element of Ω is defined for $t > \tau = 0$.

In view of (3.2), we can define the "length" $|\omega|$ of ω by

$$|\omega| = \max\{-t \in \underline{\underline{Z}}: \ \omega \text{ is not defined for any } s < t\}.$$

Before defining the semigroup on Ω, we introduce another fundamental notion of dynamics: the (left) shift operator σ_Ω, defined for all $q \geq 0$ in $\underline{\underline{Z}}$ by

(3.3) $\quad \sigma_\Omega^q\colon \ \Omega \to \Omega\colon \ \omega \mapsto \sigma_\Omega^q\omega\colon \ t \to \omega(t + q).$

Note that the definition of σ_Ω is compatible with the normalization (3.2).

If $J_\omega \cap J_{\omega'}$ = empty for $\omega, \omega' \in \Omega$, we define the join of ω and ω' as the function

(3.4) $$\omega \vee \omega' = \begin{cases} \omega & \text{on } J_\omega, \\ \omega' & \text{on } J_{\omega'}. \end{cases}$$

When Ω has an additive structure, then we replace $\omega \vee \omega'$ by $\omega + \omega'$.

(3.5) $\quad$ DEFINITION. There is an associative operation $\circ\colon \ \Omega \times \Omega \to \Omega$, called concatenation, defined by

$$\circ\colon \ (\omega, \nu) \mapsto \sigma_\Omega^{|\nu|}\omega \vee \nu.$$

Note that, by (3.2) through (3.4), $\circ$ is well defined.

Note also that the asserted existence of concatenation rests on the fact that Ω is made up of functions defined over finite intervals in T. We might express the content of (3.5) also as: Ω is a semigroup with valuation, since evidently $|\omega\circ\nu| = |\omega| + |\nu|$.

In view of (3.5), it is natural to use an abbreviated notation* also for the transition function, as follows:

(3.6) $x \circ \omega = \varphi(0; -|\omega|, x, \omega)$

Now we come to an important nonclassical concept in dynamical systems, whose evolution was strongly influenced by problems in communications and automata theory: a <u>discrete-time constant input/output map</u>

(3.7) $f: \Omega \to Y: \omega \mapsto f(\omega) = y(1)$

We interpret this map as follows: $y(1)$ is the output of some system Σ (say, a digital computer) when Σ is subjected to the (finite) input sequence ω, assuming that Σ is some fixed initial equilibrium state before the application of ω. This definition automatically incorporates the notions of "discrete-time" as well as "causal" or "dynamics" (the latter because $y(t)$ is <u>not defined</u> for $t < 1$). However, (3.7) does not clearly imply "constancy" (implicitly, however, this is clear from the normalization assumption (3.2) on Ω). To make the definition more forceful, we extend f to the map

(3.8) $\bar{f}: \Omega \to \Gamma = Y \times Y \; \ldots$ (infinite cartesian product)

$: \omega \mapsto (f(\omega), f(\sigma_\Omega \omega), \ldots) = (y(1), y(2), \ldots).$

Interpretation: $\bar{f}$ gives the output sequence $\gamma = (y(1), y(2), \ldots$ of the system Σ after $t = 0$ resulting from the application of an

----

*Observe that $x \circ \omega$ is the strict analog of the notation xt customary in topological dynamics. The action of ω on x satisfies $x \circ (\omega \circ \nu) = (x \circ \omega) \circ \nu$ in view of (1.5).

input ω which stops at $t = 0$.

This definition expresses causality more forcefully and incorporates constancy, *provided* we define the (left) shift operator σ_Γ on Γ so as to be compatible with (3.3). So, for any $\tau \geqq 0$, $\tau \in \underline{\underline{Z}}$, let

$$(3.9) \qquad \sigma_\Gamma\colon\ \Sigma \to \Gamma\colon\ \gamma \mapsto \sigma_\Gamma\gamma\colon\ t \mapsto \gamma(t+\tau)$$
$$:(y(1),\ y(2),\ \ldots\ ) \mapsto (y(\tau+1),\ y(\tau+2),\ \ldots\ )$$

Note: the operator σ_Ω "appends" an undefined term at 0, the operator σ_Γ "discards" the term $y(1)$.

Now, dropping the bar over f, we adopt

(3.10) DEFINITION. *A discrete-time, constant input/output map (of some underlying dynamical system* Σ*) is any map* f *such that the following diagram*

$$\begin{tikzcd}
\Omega \arrow[r, "f"] \arrow[d, "\sigma_\Omega"'] & \Gamma \arrow[d, "\sigma_\Gamma"] \\
\Omega \arrow[r, "f"'] & \Omega
\end{tikzcd}$$

is commutative. We say that f *is linear iff it is a K-vector space homomorphism.*

It will be convenient to regard (3.10) as the *external* definition of a dynamical system, in contrast to the *internal* definition set up in Section 1.

Intuitively, we should think of f as a highly idealized kind of experimental data; namely, f incorporates all possible information that could be gained by subjecting the underlying system to experiments in which only input/output data is available. This point of view is related to experimental physics the same way as the classical notion of a dynamical system is related to Newtonian (axiomatic) physics.

The basic question which motivates much of what will follow can now be formulated as follows:

(3.11) PROBLEM OF REALIZATION. *Given only the knowledge of* f *(but of course also of* $\underline{\underline{Z}}$, Ω, *and* Γ) *how can we discover, in a mathematically consistent, rigorous, and natural way, the properties of the system* Σ *which is supposed to underlie the given input/output map* f?

This suggests immediately the following fundamental concept:

(3.12) DEFINITION. *A fixed dynamical system* Σ *(internal definition, as in Section 1) is a realization of a fixed input/output map* f_o *iff* $f_o = f_{\Sigma_o}$, *that is,* f_o *is identical with the input/output map of* Σ_o.

In view of the notations of Section 1 plus the special convention (3.6), the explicit form of the *realization condition* is simply that

(3.13) $$f_o(\omega) = \eta_{\Sigma_o}(\varphi_{\Sigma_o}(0; -|\omega|, *, \omega))$$

for all ω Ω. The symbol $*$ stands for an arbitrary equilibrium state in which Σ_o remains, by definition, until the application of ω. (Later we simply take $*$ to be 0.)

To solve the realization problem, the critical step is to induce a definition of X (of some Σ_o) from the given f_o. It is rather surprising that this step turns out to be trivial, on the abstract level. (On the concrete level, however, there are many unsolved problems in actually *computing* what X is. In Section 8, we shall solve this problem, too, but only in the linear case.) The essential idea seems to have been published first by NERODE [1958]:

(3.14) DEFINITION. *Make the concatenation semigroup* Ω *into a monoid by adjoining a neutral element* $\emptyset$ (*which is the nowhere-defined function on* $\mathbf{Z}$). *Then* $\omega \equiv_f \omega'$ (*read:* ω *is Nerode equivalent to* ω' *with respect to* f) *iff*

$$f(\omega \circ \nu) = f(\omega' \circ \nu) \quad \underline{\text{for all}} \quad \nu \in \Omega.$$

There are many intuitive, physical, historical, and technical reasons (which are scattered throughout the literature and concentrated especially strongly in KALMAN, FALB, and ARBIB [1969]) for using this as the

(3.15) MAIN DEFINITION. *The set of equivalence classes under* $\equiv_f$, *denoted as* $X_f = \{(\omega)_f\colon \omega \in \Omega\}$, *is the state set of the input/output map* f.

Let us verify immediately that (3.15) makes mathematical sense:

(3.16) PROPOSITION. <u>For each linear, constant input/output map</u> f <u>there exists a dynamical system</u> Σ_f <u>such that</u>

(a) Σ_f <u>realizes</u> f;

(b) $X_{\Sigma_f} = X_f$.

PROOF. We show how to induce Σ_f, given f. We define the state set of Σ_f by (b). Further, we define the transition function of Σ_f by

$$(3.17) \qquad x \circ \nu = (\omega)_f \circ \nu \triangleq (\omega \circ \nu)_f \quad \text{for all} \quad \nu \in \Omega,\ x \in X_f.$$

We must check that $\circ$ on the left of $\triangleq$ is well defined (note two different uses of $\circ$!), that is, independent of the representation of x as $(\omega)_f$. This follows trivially from (3.14). Now we define the <u>readout map</u> of Σ_f by

$$(3.18) \qquad \eta_{\Sigma_f}: X_f \to Y: (\omega)_f \mapsto f(\omega)(1)$$

Again, this map is well defined since we can take $\nu = \emptyset$ as a special case in (3.14). Then

$$\eta_{\Sigma_f}(x \circ \nu) = \eta_{\Sigma_f}((\omega \circ \nu)_f) = f(\omega \circ \nu),$$

and the realization condition (3.6) is verified. Hence claim (a) is correct. □

(3.19) COMMENTS. In automata theory, Σ_f is known as the <u>reduced form</u> of any system which realizes f. Clearly, any two reduced forms are isomorphic, *in the set-theoretic sense,* since the set X_f is intrinsically defined by f. (This observation is a weak version of Theorem D of the Introduction; here "uniqueness" means "modulo a permutation of the labels of elements in the set X_f".) Notice also that Σ_f *is completely reachable* since, by Definition (3.15), every element $x = (\omega)_f$ of X_f is reachable via any element ω' in the Nerode equivalence class $(\omega)_f$. As to observability of Σ_f, see Section 10.

4. MODULES INDUCED BY LINEAR INPUT/OUTPUT MAPS

We are now ready to embark on the main topics of these lectures. It is assumed that the reader is conversant with modern algebra (especially: abelian groups, commutative rings, fields, modules, the ring of polynomials in one variable,and the theory of elementary divisors), on the level of, say, VAN DER WAERDEN, LANG [1965], HU [1965] or ZARISKI and SAMUEL [1958, Vol. 1]. The material covered from here on dates from 1965 or later.

Standing assumptions until Section 10:

(4.1) *All systems* $\Sigma = (F, G, H)$ *are discrete-time, linear, constant, defined over a fixed field* K (*but not necessarily finite-dimensional*).

Our immediate objective is to provide the setup and proof for the

(4.2) FUNDAMENTAL THEOREM OF LINEAR SYSTEM THEORY. *The natural state set* X_f *associated with a discrete-time, linear, constant input-output map* f *over a fixed field* K *admits the structure of a finitely generated module over the ring* $K[z]$ *of polynomials* (*with indeterminate* z *and coefficients in* K).

(4.3) COMMENTS. Since the ring $K[z]$ will be seen to be related to the inputs to Σ, this result has a superficial resemblance to the fact that in an *arbitrary* dynamical system Σ the state set X_Σ admits the action of a semigroup, namely Ω_Σ (see (3.6) and related footnote). It turns out, however, that this action of Ω on X, which results from combining the concatenation product in Ω with the definition of states via Nerode equivalence, is incompatible with the additive structure of Ω [KALMAN, 1967, Section 3]. Our theorem asserts the existence of an entirely different kind of structure of X. This structure, that of a $K[z]$-module, is not just a consequence of dynamics, but depends critically on the additive structure on Ω and on the linearity of f. The relevant multiplication is not (noncommutative) concatenation but (commutative) convolution (because convolution is the natural product in $K[z]$); dynamics is thereby restated in such a way that the tools of commutative algebra become applicable. In a certain rather definite sense (see also Remark (4.30)), Theorem (4.2) expresses the algebraic content of the method of the Laplace transformation, especially as regards the practices developed in electrical engineering in the U.S. during the 1950's.

The proof of Theorem (4.2) consists in a long sequence of canonical constructions and the verification that everything is well defined and works as needed.

In view of (4.1) and the conventions made in Section 1, Ω may be viewed as a K-vector space and $\omega(t) = 0$ for almost all $t \in \underline{\underline{Z}}$ and all $\omega \in \Omega$. By convention (3.2), we have assumed also that $\omega(t) = 0$ for all $t > 0$. As a result, we have that:

(a) $\Omega \approx K^m[z]$ <u>as a K-vector space</u>. Let us exhibit the isomorphism explicitly as follows:

$$(4.4) \qquad \omega \approx \sum_{t \in \underline{\underline{Z}}} \omega(t) z^{-t} \in K^m[z].$$

By (3.2), the sum in (4.4) is always finite. The isomorphism obviously preserves the K-linear structure on Ω. In the sequel, we shall not distinguish sharply between ω as a function $T \to K^m$ and ω as an m-vector polynomial.

(b) Ω *is a free* $K[z]$*-module with* m *generators, that is,* $\Omega \approx K^m[z]$ *also in the* $K[z]$*-module sense*. In fact, we define the action of $K[z]$ on Ω by scalar multiplication as

$$\cdot : \; K[z] \times \Omega \to \Omega : \; (\pi, \omega) \mapsto \pi\cdot\omega$$

where

$$(4.5) \qquad \pi\cdot\omega = \begin{bmatrix} \pi\omega_1 \\ \vdots \\ \pi\omega_m \end{bmatrix} \qquad (\omega_j \in K[z], \; j = 1, \dots, m).$$

The product of π with the components of the vector ω is the product in $K[z]$. We write the scalar product on the left, to avoid any confusion with notation (3.6). It is easy to see that the module axioms are verified; Ω is obviously free, with generators

$$(4.6) \qquad e_j = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \longleftarrow \text{j-th position, } j = 1, \dots, m.$$

(c) *On* Ω *the action of the shift operator* σ_Ω *is represented by multiplication by* z. This, of course, is the main reason for introducing the isomorphism (4.4) in the first place.

(d) *Each element of* Γ *is a formal power series in* z^{-1}. In fact, (4.4) suggests viewing z^{t} as an abstract representation of $-t \in \mathbb{Z}$; hence we define

$$(4.7) \qquad \gamma \approx \sum_{t\in\mathbb{Z}} \gamma(t)z^{-t} \in K^{p}[[z^{-1}]].$$

By (3.8) and (4.1), $\gamma(t) \in K^{p}$ for each $t > 1$ and is zero (or not defined) for $t < 1$. In general the sum is taken over infinitely many nonzero terms; there is no question of convergence and the right-hand side of (4.7) is to be interpreted stictly algebraically as a formal power series. Since $\gamma(0)$ is always zero (see (3.8)), we can say also that

(e) Γ *is isomorphic to the* K-*vector subspace of* $K^{p}[[z^{-1}]]$ *(formal power series in* z^{-1} *with coefficients in* K^{p}*) consisting of all power series with* 0 *first term*.

The first nontrivial construction is the following:

(f) Γ *has the structure of a* $K[z]$ *module, with scalar multiplication defined as*

$$(4.8) \qquad \cdot : \; K[z] \times \Gamma \to \Gamma: \; (\pi, \gamma) \mapsto \pi\cdot\gamma \; = \; \pi(\sigma_{\Gamma})\gamma.$$

This product may be interpreted as the ordinary product of a power series in z^{-1} by a polynomial in z, followed by the deletion of all terms containing no negative powers of z. The verification of the module axioms is straightforward.

(g) f *is a* $K[z]$ *homomorphism*. This is an immediate consequence of the fact that f = constant (see (3.10)) and that multiplication by z corresponds to the left shift operators on Ω and Γ.

(h) *The Nerode equivalence classes of* f *are isomorphic with* $\Omega/\text{kernel } f$. This is an easy but highly nontrivial lemma, connecting Nerode equivalence with the module structure on Ω. The proof is an immediate consequence of the formula

$$(4.9) \qquad \omega \circ \nu = z^{|\nu|}\omega + \nu.$$

In fact, by K-linearity of f, (4.9) implies

$$f(\omega \circ \nu) = f(\omega' \circ \nu) \quad \text{for all} \quad \nu \in \Omega$$

if and only if

$$f(z^k \cdot \omega) = f(z^k \cdot \omega') \quad \text{for all} \quad k \geqq 0 \quad \text{in} \quad \underline{\underline{Z}}.$$

The proof of Theorem (4.2) is now complete, since the last lemma identifies X_f as defined by (3.15) with the $K[z]$ quotient module $\Omega/\text{kernel } f$.

We write elements of the latter as $[\omega]_f = \omega + \text{kernel } f$; then it is clear that X_f as a $K[z]$-module is generated by $[e_1]_f, \ldots, [e_m]_f$, since Ω itself is generated by $e_1, \ldots, e_m$ (see (4.6)). Note also that the scalar product in $\Omega/\text{kernel } f$ is

$$(4.10) \qquad (\pi, [\omega]_f) \mapsto \pi \cdot [\omega]_f = [\pi \cdot \omega]_f.$$

The last product above (that in Ω) has already been defined in (4.5). The reader should verify directly that (4.10) gives a well-defined scalar product.

(4.11) REMARK. There is a strict duality in the setup used to define f. From the point of view of homological algebra [MAC LANE 1963], this duality looks as follows. Since every free module is projective, *the natural map*

$$\mu: \ \Omega \to X_f: \ \omega \mapsto [\omega]_f$$

exhibits X_f *as the image of a projective module*. On the other hand, there is a bijection between the set X_f and the set

$$\Xi_f = f(\Omega) \subset \Gamma.$$

Ξ_f is clearly a $K[z]$-submodule of Γ (with $z \cdot f(\omega) = f(z \cdot \omega)$), and so X_f and Ξ_f are isomorphic also as $K[z]$-modules. It is known that Γ is an injective module [MAC LANE 1963, page 95, Exercise 2] So *the natural map* $X_f \to \Xi_f$: $[\omega]_f \mapsto f(\omega)$ *exhibits* X_f *as a submodule of an injective module*. This fact is basic in the construction of the "transfer function" associated with f (Section 7), but its full implications are not yet understood at present.

There is an easy counterpart of Theorem (4.2) which concerns a dynamical system given in "internal" form:

(4.12) PROPOSITION. *The state set* X_Σ *of every discrete-time, finite-dimensional, linear, constant dynamical system* $\Sigma = (F, G, -)$ *admits the structure of a* $K[z]$-*module*.

PROOF. By definition (see (1.10)), $X = K^n$ is already a K-vector space. We make it into a $K[z]$-module by defining

(4.13) $\quad \cdot : K[z] \times K^n \to K^n: \quad (\pi, x) \mapsto \pi(F)x.$ □

(4.14) COMMENT. The construction used in the proof of (4.12) is the classical trick of studying the properties of a fixed linear map $F: K^n \to K^n$ via the $K[z]$-module structure that F induces on K^n by (4.13). In view of the canonical construction of Σ_f provided by Proposition (3.16), the state set X can be treated as a $K[z]$-module irrespective as to whether X is constructed from f $(X = X_f)$ or given a priori as part of the specification of Σ $(X = X_\Sigma)$. Thus the $K[z]$-module structure on X is a nice way of uniting the "external" and the "internal" definitions of a dynamical system. Henceforth we shall talk about a (discrete-time, linear, constant dynamical) system Σ somewhat imprecisely via properties of its associated $K[z]$-module X_Σ.

We shall now give some examples of using module-theoretic language to express standard facts encountered before.

(4.15) PROPOSITION. _If_ X _is the state-module of_ Σ, _the map_ F_Σ _is given by_ $X \to X: x \mapsto z \cdot x$.

PROOF. This is obvious from (4.13) if $X = X_\Sigma$. If $X = X_f = X_{\Sigma_f}$, then we find that, by (1.17),

$$\begin{aligned} x(1) &= Fx(0) + G\omega(0), \\ &= F[\xi]_f + G\omega(0); \end{aligned}$$

since $x(0)$ results from input ξ, $x(1)$ results from input $z \cdot \xi + \omega(0)$

and we get

$$\begin{aligned} &\doteq [z\cdot\xi + \omega(0)]_f, \\ &= z\cdot[\xi]_f + [\omega(0)]_f, \\ &= z\cdot[\xi]_f + G\omega(0). \end{aligned}$$

So the assertion is again verified. □

Now we can replace Proposition (2.14) by the much more elegant

(4.16) PROPOSITION. A system $\Sigma = (F, G, -)$ is completely reachable if and only if the columns of G generate X_Σ.

PROOF. The claim is that complete reachability is equivalent to the fact that every element $x \in X_\Sigma$ is expressible as

$$x = \sum_{j=1}^{m} \pi_j g_j, \quad \pi_j \in K[z], \quad G = [g_1, \ldots, g_m].$$

In view of (4.15), this is the same as requiring that x be expressible as

$$x = \sum_{j=1}^{m} \pi_j(F) g_j;$$

this last condition is equivalent to complete reachability by (2.14). □

(4.17) COROLLARY. The reachable states of Σ are precisely those of the submodule of X_Σ generated by (the columns of) G.

(4.18) REMARK. The statement that "Σ is not completely reachable" simply means that X is not generated by those vectors which make up the matrix G in the specification of the input side of the system Σ.

It does not follow that X cannot be finitely generated by some other vectors. In fact, to avoid unnecessary generality, we shall henceforth assume that

X is always finitely generated over $K[z]$.

From the system-theoretic point of view, the case when we need infinitely many generators, that is, infinitely many input channels, seems rather bizzare at present.

(4.19) PROPOSITION. The system X_f is completely reachable.

PROOF. Obvious from the notation: a state $x = [\xi]_f$ is reached by $\xi \in \Omega$. □

(4.20) PROPOSITION. The system X_f is completely observable.

PROOF. Obvious from Lemma (h) above: $\eta([\omega]_f) = f(\omega) = 0$ iff $\omega \in [0]_f$, which says that the only unobservable state of X_f is $0 \in X_f$. □

Let us generalize the last result to obtain a module-theoretic criterion for complete observability. There are two technically different ways of doing this. The first depends on the observation that the "dual" of a submodule (see Corollary (4.17)) is a quotient module. The second defines observability via the "dual" system $(F', H', -)$ associated with $(F, -, H)$.

Consider a dynamical system $\Sigma = (F, -, H)$ and the corresponding $K[z]$-module X_Σ and K-homomorphism $H: X_\Sigma \to Y = K^p$. We can extend H

to a K[z]-homomorphism $\bar{H}$ (look back at (2.8)) by setting

$$\begin{aligned} \bar{H}:\ & X_\Sigma \to \Gamma \\ & x \mapsto (Hx,\ H(z\cdot x),\ H(z^2\cdot x),\ \dots\ ). \end{aligned}$$

From Definition (2.19) we see that no nonzero element of the quotient module $X_\Sigma/\text{kernel}\ \bar{H}$ is unobservable. Hence, by abuse of language, we can say that $X_\Sigma/\text{kernel}\ \bar{H}$ is the module of observable states of Σ. Thus we arrive at phrasing the counterparts of (4.16-17) in the following language:

(4.21) PROPOSITION. *A system* $\Sigma = (F, -, H)$ *is completely observable if and only if the quotient module* $X_\Sigma/\text{kernel}\ \bar{H}$ *is isomorphic with* X_Σ.

(4.22) COROLLARY. *The observable states of* Σ *are to be identified with the elements of the quotient module* $X_\Sigma/\text{kernel}\ \bar{H}$.

(4.23) TERMINOLOGY. The preceding considerations suggest viewing a system Σ as essentially the same "thing" as a module X. Strictly speaking, however, knowing $\Sigma = (F, G, H)$ gives us not only $X_\Sigma = X_F$ (see (4.13)) but also a quotient module X_Σ^o (over kernel $\bar{H}$) of a submodule (that generated by G) of X_F, that is

$$X_\Sigma^o = K[z]G/\text{kernel}\ \bar{H}.$$

If $X_\Sigma^o \approx X_\Sigma$ we say that X_Σ is *canonical* (relative to the given G, H).

To be more precise, let us observe the following stronger version of (4.19-20):

(4.24) CORRESPONDENCE THEOREM. *There is a bijective correspondence between* $K[z]$-*homomorphisms* $f: \Omega \to \Gamma$ *and the equivalence class of completely reachable and completely observable systems* Σ *modulo a basis change in* X_Σ.

Detailed discussion of this result is postponed until Section 7.

A stricter observation of the "duality principle" leads to

(4.25) DEFINITION. *The* K-*linear dual of* $\Sigma = (F, G, H)$ *is* $\Sigma* = (F', H', G')$ (' = *matrix transposition*). *The states of* $\Sigma*$ *are called costates of* Σ.

The following fact is an immediate consequence of this definition:

(4.26) PROPOSITION. *The state set* $X_{\Sigma*}$ *of* $\Sigma*$ *may be given the structure of* $K[z^{-1}]$ *module, as follows:* (i) *as a vector space* $X_{\Sigma*}$ *is the dual of* X_Σ *regarded as a* K-*vector space,* (ii) *the scalar product in* $X_{\Sigma*}$ *is defined by*

$$(z^{-1} \cdot x*)(x) = x*(Fx).$$

(4.26A) REMARK. We *cannot* define $X_{\Sigma*}$ as $\mathrm{Hom}_{K[z]}(X_\Sigma, K[z])$ equal to $K[z]$-linear dual of X_Σ, because every^torsion module M over an integral domain D has a trivial D-dual. However, the reader can verify (using the ideas to be developed in Section 6) that $X_{\Sigma*}$ defined above is isomorphic with $\mathrm{Hom}_{K[z]}(X_\Sigma, K(z)/K[z])$. See BOURBAKI [*Algèbre*, Chapter 7 (2^e éd.), Section 4, No. 8].

Now we verify easily the following dual statements of (4.16-17):

(4.27) PROPOSITION. *A system* $\Sigma = (F, -, H)$ *is completely observable if and only if* H' *generates* $X_{\Sigma*}$.

(4.28) COROLLARY. *The observable* CO*states of* $\Sigma*$ *are precisely the reachable states of* $\Sigma*$, *that is, those of the submodule of* $X_{\Sigma*}$ *generated by* H'.

We have eliminated the abuse of language incurred by talking about "observable states" through introduction of the new notion of "observable COstates". The full explication of why this is necessary (as well as natural) is postponed until Section 10.

The preceding simple facts depend only on the notion of a module and are immediate once we recognize the fact that F may be eliminated from statements such as (2.8) by passing to the module induced by F via (4.13). But module theory yields many other, less obvious results as well, which derive mainly from the fact that $K[z]$ is a principal-ideal domain.

We recall: an element m of an R-module M (R = arbitrary commutative ring) has *torsion* iff there is a $r \in R$ such that $r \cdot m = 0$. If this is not the case, m is *free*. Similarly, M is said to be a *torsion module* iff every element of M has torsion. M is a *free module* if no nonzero element has torsion. If $L \subset M$ is any subset of M, the annihilator A_L of L is the set

$$A_L = \{r: r \cdot \ell = 0 \text{ for all } \ell \in L\};$$

it follows immediately that A_L is an ideal in R. Note also that the statement "M is a torsion module" does not imply in general that A_L is nontrivial, that is, $A_L \neq 0$. (Counterexample: take an M which is not finitely generated.)

Coupling these notions with the special fact that, for us, $R = K[z]$, we get a number of interesting system-theoretic results:

(4.29) PROPOSITION. Σ *is finite-dimensional if and only if* X_Σ *is a torsion* $K[z]$-*module*.

COROLLARY. *If* X_Σ *is free,* Σ *is infinite dimensional*.

PROOF. We recall that "Σ = finite-dimensional" is defined to be "X_Σ = finite-dimensional as a K-vector space". See (1.18).

Sufficiency. By assumption X is finitely generated by, say, q nonzero elements $x_1, \ldots, x_q$ of X_Σ (which are not necessarily the columns of G). Hence

$$A_X = A_{x_1} \cap \ldots \cap A_{x_q}$$

Since $K[z]$ is a principal-ideal domain, each of the A_{x_j} is a principal ideal, say, $\gamma_j K[z]$ with $\gamma_j \in K[z]$. If X_Σ is a torsion module, then $\deg \gamma_j = n_j > 0$ for all $j = 1, \ldots, q$. For otherwise γ_j is either zero (and then x_j is free, which is a contradiction) or a unit which implies $x_j = 0$ contr y to assumption. Hence we can replace each expression

$$x = \sum_{j=1}^{q} \pi_j \cdot x_j, \quad \pi_j \in K[z]$$

by the simpler one

$$x = \sum_{j=1}^{q} [\pi_j \ (\mathrm{mod}\ \gamma_j)]\cdot x_j,$$

which shows that X_Σ, as a K-module, is generated by the finite set

$$x_1, \ z\cdot x_1, \ \ldots, \ z^{n_1 - 1}\cdot x_1, \ x_2, \ \ldots, \ x_q.$$

Necessity. Let ψ_F be the minimal polynomial of the map $F\colon \ x \mapsto z\cdot x$. If X_Σ is finite-dimensional as a K-module, $\deg \psi_F > 0$. This means (by the usual definition of the minimal polynomial in matrix theory or more generally in linear algebra) that ψ_F annihilates every $x \in X_\Sigma$ so that X_Σ is a torsion $K[z]$-module. □

Notice, from the second half of the proof, that the notion of a minimal polynomial can be extended from K-linear algebra to $K[z]$-modules. In fact, the same argument gives us also the well-known

(4.30) PROPOSITION. *Every finitely generated torsion module* M *over a principal-ideal domain* R *has a nontrivial minimal p ynomial* ψ_M *given by* $A_M = \psi_M R$.

(4.31) COROLLARY. *If a* $K[z]$-*module* X *is finitely generated with* q *generators and minimal polynomial* ψ_X, *then*

$$\dim X \ \text{(as K-vector space)} \ \leqq \ q\cdot \deg \psi_X.$$

(4.32) REMARK. The fact that Σ_f is completely reachable and is therefore generated by m vectors allows us to estimate the dimension of Σ_f by (4.31) knowing only $\deg \psi_{X_f}$ but without having computed

X_f itself. (Knowing X_f explicitly means knowing $F\colon x \mapsto z\cdot x$, etc.) In other words, the module-theoretic setup considerably enhances the content of Proposition (3.16). Guided by these observations, we shall develop in Section 8 explicit algorithms for calculating $\dim \Sigma_f$ directly from f without first having to compute F.

(4.33) PROPOSITION. *If X_Σ is a free $K[z]$-module, no state of Σ can be simultaneously reachable and controllable.*

PROOF. We recall that "X_Σ = free" means that X_Σ is (isomorphic to) a finite sum of copies of $K[z]$. Suppose for simplicity that $X_\Sigma = K[z]$. Then x = reachable means that $x = \xi\cdot 1$ for some $\xi \in K[z]$. Similarly, x = controllable means that $z^{|\omega|}\cdot x + \omega\cdot 1 = 0$ for some $\omega \in K[z]$. Hence if x has both properties,

$$(z^{|\omega|}\xi + \omega)\cdot 1 \;=\; (\xi\circ\omega)\cdot 1 \;=\; 0.$$

This shows that 1 is annihilated by $\xi\circ\omega$, the input ξ followed by ω, which contradicts the assumption that X_Σ is free. □

The most important consequence of Theorem (4.2) is due to the fact that through it we can apply to linear dynamical systems the well-known

(4.34) FUNDAMENTAL STRUCTURE THEOREM FOR FINITELY GENERATED MODULES OVER A PRINCIPAL IDEAL DOMAIN R (Invariant Factor Theorem for Modules). *Every such module M with m generators is isomorphic to*

(4.35) $R/\psi_1 R \oplus \ldots \oplus R/\psi_r R \oplus R^s$

where the $R/\psi_i R$ *are quotient rings of* R *viewed as modules over* R, *the* ψ_i (*called the invariant factors of* M) *are uniquely determined by* M *up to units in* R, $\psi_i | \psi_{i-1}$, $i = 2, \ldots, q$, *and, as usual,* R^s *denotes the free* R-*module with* s *generators; finally,* $r + s \leqq m$.

Various proofs of this theorem are referenced in KALMAN, FALB, and ARBIB [1969, page 270], and one is given later in Section 6.

Note: The divisibility conditions imply that M *is a torsion module iff* $s = 0$ and then $\psi_M = \psi_1$.

One important consequence of this theorem (others in Section 7) is that it gives us the most general situation when X_Σ is not a torsion module Σ. For instance, combining (4.33) with (4.34), we get

(4.36) PROPOSITION. *A system cannot be simultaneously completely reachable and completely controllable if its* $K[z]$-*module* X *has any* ∞-*dimensional components (i.e.,* $s > 0$ *in* (4.35)).

(4.37) REMARK. Although our entire development in this section may be regarded as a deep examination of Proposition (2.14), most of our comments apply equally well to (2.7), since both statements rest on the *same* algebraic condition (2.8). In fact, the only remaining thing to be "algebraized" is the notion of "continuous-time". We shall not do this here. Once this last step is taken, the algebraization of the Laplace transform (as related to *ordinary* linear differential equations) will be complete.

5. CYCLICITY AND RELATED QUESTIONS

We recall that an R-module M (R = arbitrary ring) is cyclic iff there is an element $m \in M$ such that $M = Rm$. [It would be better to say that such a module is monogenic: generated by one element m.]

If M is cyclic, the map $R \to M: r \mapsto r \cdot m$ is an epimorphism and has kernel A_m, the annihilating ideal of m. This plus the homomorphism theorem gives the well-known

(5.1) PROPOSITION. Every cyclic R-module M with generator m is isomorphic with the quotient ring R/A_m viewed as an R-module.

This result is much more interesting when, as in our case, R is not only commutative and a principal-ideal domain, but specifically the polynomial ring $K[z]$.

So let X be a cyclic $K[z]$-module with generator g and let $A_g = \psi_g K[z]$, where ψ_g is the minimal or annihilating polynomial of g. By commutativity and cyclicity, $A_g = A_X$. Hence ψ_g is a minimal polynomial also for X. Write $\psi_g = \psi_X = \psi$. In view of (5.1), $X \approx K[z]/\psi K[z]$. Let us recall some features of the ring $K[z]/\psi K[z]$:

(i) Its elements are the residue classes of polynomials $\pi \pmod{\psi}$, $\pi \in K[z]$. Write these as $[\pi]$ or $[\pi]_\psi$. Multiplication is defined as $[\pi] \cdot [\sigma] = [\pi\sigma]$.

(ii) Each $[\pi]$ is either a unit or a divisor of zero. In fact, $[\pi]$ is a unit iff (π, ψ) = greatest common divisor of π, ψ is a unit in $K[z]$ (that is, $(\pi, \psi) \in K$). Then

$$\sigma\pi + \tau\psi = 1 \quad (\sigma, \tau \in K[z])$$

so that $[\sigma]$ is the inverse of $[\pi]$. On the other hand, if $(\pi, \psi) = 0 \neq$ unit in $K[z]$, then both $[\pi]$ and $[\psi/\theta]$ are zero divisors since $[\pi]\cdot[\psi/\theta] = [(\pi/\theta)\psi] = 0$.

(iii) If ψ is a prime in $K[z]$ (that is, an irreducible polynomial with respect to coefficients over the ground field K), then by (ii) $K[z]/\psi K[z]$ is a field. This is a very standard construction in algebraic number theory.

Since it is awkward to compute with equivalence classes $[\pi]$, we shall often prefer to work with the standard representative of $[\pi]$, namely a polynomial $\tilde{\pi}$ of least degree in $[\pi]$. $\tilde{\pi}$ is uniquely determined by $[\pi]$ and the condition $\deg \tilde{\pi} < \deg \psi$. Henceforth $\tilde{}$ will always be used in this sense.

The next two assertions are immediate:

(5.2) PROPOSITION. $K[z]/\psi K[z]$ *as a* K-*vector space is isomorphic to the* K-*vector space* $\textcircled{P}^{(n)} = \{\tilde{\xi} \in K[z]: \deg \tilde{\xi} < n = \deg \psi\}$. $K[z]/\psi K[z]$ *is also isomorphic to* $\textcircled{P}^{(n)}$ *as a* $K[z]$-*module, provided we define the scalar product in* $\textcircled{P}^{(n)}$ *by* $(\pi\cdot\tilde{\xi}) \mapsto \widetilde{\pi\tilde{\xi}}$.

(5.3) PROPOSITION. *If* X_Σ *is cyclic with minimal polynomial* ψ, *then* $\dim \Sigma = \deg \psi$.

Looking back at Theorem (4.34), we see that the most general $K[z]$-module is a direct sum of cyclic $K[z]$-modules. By combining (5.3) and (4.34) and using the fact that dimension is additive under direct summing, we can replace (4.31) by the following exact result:

(5.4) PROPOSITION. *If* X_Σ *is a torsion module with invariant factors* $\psi_1, \ldots, \psi_q$ *then*

$$\dim \Sigma = \deg \psi_1 + \ldots + \deg \psi_q.$$

A simple but highly useful consequence of cyclicity is the so-called *control canonical form* [KALMAN, FALB, and ARBIB, 1969, page 44] for a completely reachable pair (F, g) where g is an $n \times 1$ matrix. We shall now proceed to deduce this result.

Observe first that "(F, g) completely reachable" is equivalent to "g generates X_F, the module induced by F via (4.13)." Let

$$\begin{aligned} \chi_F(z) &= \det(zI - F), \\ &= z^n + \alpha_1 z^{n-1} + \ldots + \alpha_n, \quad \alpha_1 \in K; \end{aligned}$$

then χ_F is the characteristic (and also the) minimal polynomial for X_F. [This is a well-known fact of module theory. See for example KALMAN, FALB, and ARBIB [1969, Chapter 10, Section 7] for detailed discussion.] As in KALMAN [1962], consider the vectors

$$
(5.5)\quad \begin{cases} e_n = g = 1\cdot g = \chi_F^{(1)}(z)\cdot g, \\ e_{n-1} = z\cdot g + \alpha_1\cdot g = \chi_F^{(2)}(z)\cdot g, \\ \vdots \\ e_1 = z^{n-1}\cdot g + a_1 z^{n-2}\cdot g + \ldots + \alpha_{n-1}\cdot g = \chi_F^{(n)}(z)\cdot g \end{cases}
$$

in X_F. [For consistency, $\chi_F^{(n+1)}(z) = \chi_F(z)$.] These vectors are easily seen to be linearly independent over K. They generate X_F since $X_F \approx \textcircled{P}^{(n)}$ as a K-vector space (Proposition (5.2)). Hence $e_1, \ldots, e_n$ are a basis for X_F as a K-vector space. With respect to this basis, the K-homomorphism

$$z\colon K^n \to K^n\colon\ x \mapsto z\cdot x$$

is represented by the matrix

$$
(5.6)\qquad F = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 0 & 1 \\ -\alpha_r & -\alpha_{n-1} & -\alpha_{n-2} & \ldots & -\alpha_2 & -\alpha_1 \end{bmatrix}
$$

[This is proved by direct computation. In particular, it is necessary to use the fact that

$$\begin{aligned} z \cdot e_1 &= z\chi_F^{(n)}(z) \cdot g, \\ &= (\chi_F(z) - \alpha_n) \cdot g, \\ &= - \alpha_n \cdot e_n \cdot] \end{aligned}$$

Note that the last row of F in (5.6) consists of the coefficients of χ_F. By definition, $g = e_n$. Hence g as a column vector in K^n has the representation

$$(5.7) \qquad g = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

Conversely, suppose (F, g) have the matrix representation (5.6-7) with respect to some basis in K^n. Then (by direct computation) the rank condition (2.8) is satisfied and therefore (F, g) is completely reachable in both the continuous-time and discrete-time cases (Propositions (2.7) and (2.16)).

We have now proved:

(5.8) PROPOSITION. *The pair* (F, g) *is completely reachable if and only if there is a basis relative to which* F *is given by* (5.6) *and* g *by* (5.7).

(5.9) COROLLARY. *Given an arbitrary* n-th *degree polynomial* $\lambda(z) = z^n + \beta_1 z^{n-1} + \dots + \beta_n$ *in* $K[z]$, $K =$ *arbitrary field. There exists an* n-*vector* ℓ *such that* $\lambda = \chi_{F-g\ell'}$ *if and only if the pair* (F, g) *is completely reachable.*

PROOF. Suppose that (F, g) is completely reachable. With respect to the same basis (5.5) which exhibits the canonical forms (5.6-7), define

$$(5.10) \qquad \ell = \begin{bmatrix} \beta_n - \alpha_n \\ \vdots \\ \beta_1 - \alpha_1 \end{bmatrix}.$$

Then verify by direct computation that $\lambda = \chi_{F-g\ell'}$.

Conversely, suppose that (F, g) is not completely reachable. Then, recalling Proposition (2.12) (which is an algebraic consequence of (2.8) and hence equally valid for both continuous-time and discrete-time), $\dim X_2 > 0$ and so is also $\deg \chi_{F_{22}}$. Since X_1 is an F-invariant subspace of $X = K^n$, the polynomial $\chi_{F_{11}}$ is independent of the choice of basis in K^n and the same is true then also for $\chi_{F_{22}} = \chi_F/\chi_{F_{11}}$. (In particular, $\chi_{F_{22}}$ does not depend on the arbitrary choice of X_2 in satisfying the condition $X = X_1 \oplus X_2$.) In view of (2.12), we have for all n-vectors ℓ,

$$\chi_{F-g\ell'} = \chi_{F_{11}-g_1\ell_1'} \cdot \chi_{F_{22}}, \quad \deg \chi_{F_{22}} > 0.$$

This contradicts the claim that $\lambda = \chi_{F-g\ell'}$ is true for any λ with suitable choice of ℓ. □

In view of the importance of this last result, we shall rephrase it in purely module theoretic terms:

(5.11) THEOREM. *Let* K *be an arbitrary field and* X *a cyclic* K[z]-*module with generator* g *and minimal polynomial* χ *of degree* n. *There is a bijection between* n-th *degree polynomials* $\lambda(z) = z^n + \beta_1 z^{n-1} + \ldots + \beta_n$ *in* K[z] *and* K-*homomorphisms* $\ell: K^n \to K^n: \chi^{(j)}\cdot g \mapsto \ell_j \cdot g$ $(j = 1, \ldots, n$ *and* $\chi^{(j)}$ *defined as in* (5.5)) *such that* λ *is the minimal polynomial for the new module structure induced on* X *by the map* $z_*: x \mapsto z\cdot x - \ell(x)$.

Note that in (5.11) $\ell(x)$ corresponds to $g\ell'x$ in (5.10).

The map ℓ in (5.11) defines a *control law* for the system $\Sigma = (F, g, -)$ corresponding to the module X. The passage from z to z_* is the module-theoretic form of the well-known open-loop to closed-loop transformation used in classical linear control theory.

PROOF. Since the vectors $\chi^{(1)}\cdot g, \ldots, \chi^{(n)}\cdot g$ form a basis for K^n, ℓ is clearly a well-defined K-homomorphism. We treat ℓ formally as an element of K[z] (that is, an operator on X is a K-vector space), by writing $\ell\cdot x = \ell(\tilde{\xi}\cdot g)$, where $\tilde{\xi}$ represents the equivalence class $[\xi] = \{\xi: \xi\cdot g = x\}$. Unless identically zero, ℓ is never a K[z]-homomorphism and therefore ℓ does not commute with nonunits in K[z].

Define $\ell_j = \beta_j - \alpha_j$, $j = 1, \ldots, n$. We prove first that this choice of ℓ implies $\lambda^{(j)}(z - \ell) = \chi^{(j)}(z)$ for $j = 1, \ldots, n + 1$. Use induction on j. By definition, $\lambda^{(1)}(z - \ell) = \chi^{(1)}(z)$. In the general case,

$$\begin{aligned}
\lambda^{(j+1)}(z-\ell)\cdot g &= [(z-\ell)\lambda^{(j)}(z-\ell)+\beta_j]\cdot g && \text{(def. of } \lambda^{(j+1)}\text{),}\\
&= [(z-\ell)\chi^{(j)}(z)+\beta_j]\cdot g && \text{(inductive hypothesis),}\\
&= [z\chi^{(j)}(z)+\beta_j-\ell_j]\cdot g && \text{(def. of } \ell\text{),}\\
&= [z\chi^{(j)}(z)+\alpha_j]\cdot g && \text{(def. of } \ell_j\text{),}\\
&= \chi^{(j+1)}(z)\cdot g && \text{(def. of } \chi^{(j+1)}\text{).}
\end{aligned}$$

It follows (case $j = n + 1$) that λ annihilates X regarded as a $K[z_*]$-module. On the other hand, the $\lambda^{(1)}(z_*)\cdot g, \ldots, \lambda^{(n)}(z_*)\cdot g$ is a basis for X as a K-vector space since $\chi^{(1)}(z)\cdot g \ldots, \chi^{(n)}(z)\cdot g$ was such a basis. So X is cyclic with generator g also as a $K[z_*]$-module. Hence by Propositions (5.1-2) the annihilating ideal of g with respect to the $K[z_*]$-module structure cannot be generated by a polynomial of degree less than n, that is, λ is indeed the minimal polynomial with respect to z_*. The correspondence $\lambda \leftrightarrow \ell$ is obviously bijective. □

The proof immediately implies the following

(5.12) COROLLARY. *Let* $x = \tilde{\xi}\cdot g$ *be any element of* X *viewed as a* $K[z]$*-module. Then* x *has the representation* $\tilde{\xi}_*\cdot g$ *with respect to the* $K[z_*]$*-module structure on* X, *where* ξ *and* ξ_* *are related as*

$$\xi(z) = \sum_{j=1}^{n} \xi_j \chi^{(j)}(z)\cdot g$$

$$\xi_*(z_*) = \sum_{j=1}^{n} \xi_j \chi_*^{(j)}(z_*) g.$$

So the open-loop/closed-loop transformation is essentially a change in the canonical basis, provided X is cyclic.

It is interesting that the $\chi^{(j)}$ have long been known in Algebra (they are related to the Tschirnhausen transformation discussed extensively by WEBER [1898, §46, 54, 74, 85, 96]), but their present (very natural) use in module theory seems to be new.

**Theorem (5.11) may be viewed as the central special case of Theorem A of the Introduction. Let us restate the latter in precise form as follows:

(5.13) THEOREM. *Given an arbitrary* n-th *degree polynomial* $\lambda(z) = z^n + \beta_1 z^{n-1} + \ldots + \beta_n$ *in* $K[z]$, K = *arbitrary field. There exists an* $n \times m$ *matrix* L *over* K *such that* $\chi_{F-GL'} = \lambda$ *if and only if* (F, G) *is completely reachable.*

For some time, this result had the status of a well-known folk theorem, considered to be a straightforward consequence of (5.9). The latter has been discovered independently by many people. (I first heard of it in 1958, proposed as a conjecture by J. E. Bertram and proved soon afterwards by the so-called root-locus method.) Indeed, the passage from (5.11) to (5.13) is primarily a technical problem. A proof of (5.13) was given by LANGENHOP [1964] and subsequently simplified by WONHAM [1967]. The first proof was (unnecessarily) very long, but the second proof is also unsatisfactory; since it depends on arguments using a splitting field of K

**The material between these marks was added after the Summer School.

and fail when K is a finite field. We shall use this situation as an excuse to illustrate the power of the module-theoretic approach and to give a proof of (5.13) valid for arbitrary fields.

The procedure of LANGENHOP and WONHAM rests on the following fact, of which we give a module-theoretic proof:

(5.14) LEMMA. *Let* K *be an arbitrary but infinite field. Let* F *be cyclic* and* (F, G) *completely reachable. Then there is an* m*-vector* $a \in K^m$ *such that* (F, Ga) *is also completely reachable*.

We begin with a simple remark, which is also useful in reducing the proof of (5.13) to Lemma (5.18).

(5.15) SUBLEMMA. *Every submodule of a cyclic module over a principal-ideal domain is cyclic*.

PROOF OF (5.14). We use induction on m. The case $m = 1$ is trivial. The general case amounts to the following. Consider the submodule Y of $X = X_F$ generated by the columns $g_1, \dots, g_{m-1}$ of G. In view of (5.15), Y is cyclic. By the inductive hypothesis, we are given the existence of a cyclic generator of Y of the form $g_y = \alpha_i g_1 + \dots + \alpha_{m-1} \cdot g_{m-1}$, $\alpha_i \in K$. We must prove: for suitable $\alpha, \beta \in K$ the vector $\alpha \cdot g_Y + \beta \cdot g_m$ is a cyclic generator for X.

----

*Of course, this means that the $K[z]$-module X_F (see (4.13)) is cyclic.

By hypothesis, X has an (abstract) cyclic generator g_X. By cyclicity we have the representations

$$g_Y = \eta \cdot g_X \text{ and } g_m = \mu \cdot g_X, \quad \eta, \mu \in K[z].$$

Hence our problem is reduced to proving the following: for suitable $\alpha, \beta \in K$ the polynomial $\alpha\eta + \beta\mu$ is a unit in $K[z]/\chi_F K[z]$. This, in turn, is equivalent to proving

$$(5.16) \qquad \alpha\eta + \beta\mu \not\equiv 0 \pmod{\theta_i} \quad i = 1, \dots, r$$

where $\theta_1, \dots, \theta_r$ in $K[z]$ are the unique prime factors of χ_F. Let $\tilde{}$ mean the representative of least degree of equivalence classes mod θ_i. Then no pair $(\tilde{\eta}_i, \tilde{\mu}_i)$, $i = 1, \dots, r$ can be zero. For if one is, then $\theta_i | (\chi_F, \eta, \mu)$, that is, χ_F/θ_i annihilates the submodule $X' = K[z]g_Y + K[z]g_m$, whence X' is a proper submodule of X, contradicting the fact that (F, G) is completely reachable. If all the $\tilde{\mu}_i$ are zero, then every $\tilde{\eta}_i \neq 0$, so η is a unit in $K[z]/\chi_F K[z]$, and g_Y is already a cyclic generator. So let $\alpha = 1$. Then the condition $\tilde{\eta}_i + \tilde{\beta}\mu_i = 0$ eliminates at most r values of β from consideration. Since K is infinite by hypothesis, there are always some β which satisfy (5.16). □

An essential part of the lemma is the stipulation that $a \in K^m$. The hypothesis "F = cyclic + (F, G) = completely reachable" means that

$$g_X = \alpha_1 g_1 + \dots + \alpha_m g_m, \quad \alpha_i \in K[z];$$

that is, the lemma is trivially true for some $a \in K^m[z]$ since $g_X = Ga$. But since we want $a \in K$, there must be interaction between vector-space structure and module structure, and for this reason the lemma is nontrivial. As a matter of fact, the lemma is false when K = finite field. The simplest counterexample is provided when (5.12) rules out a *single* nonzero value of β, thereby ruling out *all* β.

(5.17) COUNTEREXAMPLE. Let $K = \mathbb{Z}/2\mathbb{Z}$, that is, the ring of integers modulo the prime ideal $2\mathbb{Z}$. Consider

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \qquad G = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

Notice that $X_F = X_1 \oplus X_2 \oplus X_3$ (as a $K[z]$-module), where the minimal polynomials of the direct summands are

$$\begin{aligned} \chi_1(z) &= z^2 + z + 1, \\ \chi_2(z) &= z^2, \\ \chi_3(z) &= z + 1. \end{aligned}$$

All these factors are relatively prime, $(\chi_1, \chi_2, \chi_3) = 1$, hence X is cyclic. Notice also that g_1 generates $X_1 \oplus X_3$ while g_2 generates $X_2 \oplus X_3$. A cyclic generator for X is

$$g_X = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

A simple calculation gives

$$g_1 = z^3 \cdot g_X, \quad g_2 = (z^4 + z^2 + 1) \cdot g_X.$$

Conditions (5.16) are here

$$\alpha \cdot 1 + \beta \cdot 0 \neq 0 \pmod{X_1},$$
$$\alpha \cdot 0 + \beta \cdot 1 \neq 0 \pmod{X_2},$$
$$\alpha \cdot 1 + \beta \cdot 1 \neq 0 \pmod{X_3}.$$

These conditions have no solution in $\underline{\underline{Z}}/2\underline{\underline{Z}}$.

At this point, the following is the situation concerning Theorem (5.13):

(1) Its counterpart, Theorem A of the Introduction, was claimed to be true in the continuous-time case under the hypc'.hesis of complete controllability.

(2) In the discrete-time case (5.13) with the preceding hypothesis Theorem A is false, because of the counterexample: the pair (F = nilpotent, $G = 0$) is completely controllable, but evidently X_{F-GL}, is independent of L. However, in view of (5.11), Theorem (5.13) might be true also in the discrete-time case if "complete controllability" is replaced by "complete reachability", this modification being immaterial in the continuous-time case.

(3) Because of (5.17), we might expect that a theorem like (5.13) is false for an <u>arbitrary</u> field K.

(4) If our general claim that reachability properties are reflected in module-theoretic properties is true, then (5.13) <u>should hold without assumptions concerning</u> K, because the principal module-theoretic fact, that $K[z]$ = principal ideal domain, is independent of the specific choice of K.

We now proceed to establish Theorem (5.13). That is, special hypotheses on K will turn out to be irrelevant.

PROOF OF (5.13). Necessity is proved exactly as in (5.8). Sufficiency will follow by induction on m, once we have proved it in the special case $m = 2$:

(5.18) LEMMA. <u>Let</u> K <u>be an arbitrary field and let</u> X <u>be a</u> $K[z]$-<u>module generated by</u> g_1, g_2. <u>There is a</u> K-<u>homomorphism</u> ℓ (<u>of the type defined in</u> (5.11)) <u>such that if</u> $z_* = z - \ell$ <u>induces a</u> $K[z_*]$-<u>module structure on</u> X <u>then</u> X <u>is cyclic with respect to this structure and is generated by either</u> $g_1 + g_2$ <u>or</u> g_2.

PROOF. Let $Y = K[z]g_1$ and $Z = K[z]g_2$.

<u>Case 1</u>. $Y \cap Z = 0$, that is, $X = Y \oplus Z$. In (5.11) take an ℓ such that $\ell(x) = 0$ for all $x \in Z$. Replacing z by $z_* = z - \ell$ will change the $K[z]$-module structure on Y but preserve that on Z. Further, choose ℓ so that the new minimal polynomial λ on Y is prime to the unchanged minimal polynomial $\chi_{F_Z} = \chi$ on Z. Thus there exist polynomials ν, σ such that $\nu\lambda + \sigma\chi = 1$. By hypothesis, every $x \in X$ has the representation

$$x = y + z = \eta \cdot g_1 + \zeta \cdot g_2.$$

Now verify that

$$\begin{aligned} x &= (\eta\sigma\chi + \zeta\nu\lambda)\cdot(g_1 + g_2), \\ &= \eta\sigma\chi\cdot g_1 + \zeta\nu\lambda\cdot g_2, \\ &= \eta(1 - \nu\lambda)\cdot g_1 + \zeta(1 - \sigma\chi)\cdot g_2, \\ &= \eta\cdot g_1 + \zeta\cdot g_2. \end{aligned}$$

Hence $g_1 + g_2$ is indeed a cyclic generator for X as a $K[z_*]$-module.

Case 2. $Y \cap Z = W \neq 0$. Let $w \in W$. By hypothesis, there is a $\xi \in K[z]$ such that $w = \xi\cdot g_2$ and therefore, by cyclicity of Y, there is also a $\eta \in K[z]$ such that $\xi\cdot g_2 = w = \eta\cdot g_1$. Take same $w \neq 0$. Then if $\eta =$ unit (mod χ_X) we are done because $\eta^{-1}\xi\cdot g_2$ generates Y, and so $Z = X$. In the nontrivial case, $\eta \neq$ unit (mod χ_X). To show: there is a suitable new module structure on X such that $\eta_* =$ unit (mod χ_*), χ_* being the minimal polynomial of X as a $K[z_*]$-module.

The main facts we need are the following:

(5.19) SUBLEMMA. *Let* χ *be a fixed element of* $K[z]$ *with* $\deg \chi = n$, F_χ *the companion matrix of* χ *given by* (5.6), X_{F_χ} *the cyclic module induced by* F_χ, *and* g *a cyclic generator of* X_{F_χ}. *Then* $\eta \in K[z]$ *is a unit modulo* χ *if and only if* $\tilde{\eta}\cdot g$ *is also a cyclic generator of* X_{F_χ}.

PROOF. Obvious. □

(5.20) SUBLEMMA. *Same notations as in* (5.19). *Write*

$$\tilde{\eta} \;=\; \sum_{j=1}^{n} \eta_j \chi^{(j)}(z) \qquad (\chi^{(j)} \;\; \textit{defined in}\; (5.5)).$$

Then $\tilde{\eta}$ *is a unit modulo* χ *if and only if*

$$\det\,(y,\, F_\chi y,\, \ldots,\, F_\chi^{n-1} y) \;\neq\; 0, \tag{5.21}$$

where y *is the column vector*

$$y \;=\; \begin{bmatrix} \tilde{\eta}_n \\ \cdot \\ \cdot \\ \cdot \\ \tilde{\eta}_1 \end{bmatrix}. \tag{5.22}$$

PROOF. Since $\chi^{(1)}, \ldots, \chi^{(n)}$ is the basis for the K-vector space of all polynomials of degree $< n$, the n-tuple $(\tilde{\eta}_1, \ldots, \tilde{\eta}_n)$ is uniquely determined by η. By definition F_χ is the matrix representing the module operator $z\colon x \mapsto z \cdot x$ relative to the special basis $e_1, \ldots, e_n$ in X_{F_χ} given by (5.5). Similarly, using one of the module axioms, we verify that

$$\begin{aligned} \tilde{\eta} \cdot g \;&=\; \sum_{j=1}^{n} [\tilde{\eta}_j \chi^{(j)}(z)] \cdot g, \\ &=\; \sum_{j=1}^{n} \tilde{\eta}_j [\chi^{(j)}(z) \cdot g], \\ &\; \sum_{j=1}^{n} \tilde{\eta}_{n-j+1} \cdot e_j; \end{aligned}$$

in other words, the numerical vector (5.22) represents the abstract vector $\tilde{\eta} \cdot g$ in X_{F_χ} relative to the same basis $e_1, \ldots, e_n$. Recall that $\tilde{\eta}\cdot g$ generates X_{F_χ} iff $(F_\chi, \tilde{\eta}(F_\chi)g)$ is complete reachable. By (2.7) the latter condition is equivalent to (5.21). The rest follows from (5.19). □

(5.23) SUBLEMMA. *Same notations as in* (5.19) *and* (5.20). *Given any nonzero numerical* n-*vector* (5.22), *there exists a polynomial* χ *such that* (5.21) *is satisfied.*

PROOF. Let $\tilde{\eta}_r$ be the first member of the sequence of numbers $\tilde{\eta}_1, \tilde{\eta}_2, \ldots$ which is nonzero. Write

$$\chi(z) = z^n + \alpha_1 z^{n-1} + \ldots + \alpha_n,$$

and determine the first r coefficients of χ by the rule

$$\begin{bmatrix} \tilde{\eta}_r & \tilde{\eta}_{r+1} & \cdots & \tilde{\eta}_n \\ 0 & \tilde{\eta}_r & \cdots & \tilde{\eta}_{n-1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & \tilde{\eta}_r \end{bmatrix} \begin{bmatrix} \alpha_r \\ \alpha_{r+1} \\ \cdot \\ \cdot \\ \cdot \\ \alpha_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}.$$

(Since all numbers belong to a field, the required values of $\alpha_r, \ldots, \alpha_n$ exist.) Now check, by computation, that these conditions reduce the matrix in (5.21) to the direct sum of two triangular matrices, each with nonzero elements on its diagonal. □

In view of (5.12), it follows from these facts that we can always choose a new $\chi_Y = \chi_\dagger$ such that $\eta_\dagger$ = unit mod $\chi_\dagger$.

The proof of Case 2 is not yet complete, however, because we must still extend the $K[z_*]$-module structure from Y to X. This is easy. Write first $Z = W \oplus Z'$ and then $X = Y \oplus Z'$, where the direct sum is now with respect to the K-module structure of X. Extend ℓ from Y to X by setting $\ell | Z' = 0$. Now we have a new minimal polynomial χ_* defined over X Since $z_* = z_\dagger$ on Y, $\eta_* = \eta_\dagger$. By (5.12), ξ is replaced by some ξ_* such that

$$(5.24) \qquad w = \xi_* \cdot g_1 = \eta_* \cdot g_2,$$

that is, our previous representation of $w \neq 0$ in W induces a similar representation with respect to the new $K[z_*]$-module structure on X. Since η_* is a unit modulo $\chi_\dagger$, we can write

$$\sigma\eta_* = 1 + \tau\chi_\dagger, \quad \text{with} \quad \sigma, \tau \in K[z_*].$$

By (5.24), we have, with respect to the $K[z_*]$-structure,

$$\begin{aligned} (\sigma\xi_*)\cdot g_2 &= \sigma\cdot(\xi_*\cdot g_2), \\ &= \sigma\cdot(\eta_*\cdot g_1), \\ &= (1 + \tau\chi_\dagger)\cdot g_1, \\ &= g_1. \end{aligned}$$

This proves that g_2 generates both Y and Z; that is, g_2 is a cyclic generator for X endowed with the $K[z_*]$-structure. The proof of Lemma (5.18). is now complete. □

It should be clear that Theorem (5.13) is not a purely module-theoretic result, but depends on the interplay between module theory, vector-spaces, and elimination theory (via (5.21)). For instance, the fact that ℓ can be extended from Y to X, which was needed in the proof of Case 2, is a typical vector-space argument.**

There are many open (or forgotten) results concerning cyclic modules which are of interest in system theory. For instance, it is easy to show that an $n \times n$ real matrix is cyclic iff a certain polynomial $\Psi \in \underline{\underline{R}}[z_1, \ldots, z_{n^2}]$ is nonzero at F; the polynomial Ψ is roughly analogous to the polynomial det in the same ring, but, unlike in the latter case, the general form of Ψ does not seem to be known.

We must not terminate this discussion without pointing out another consequence of cyclicity which transcends the module framework. Since X = cyclic with generator g is isomorphic with $K[z]/\chi_g K[z]$, it is clear that X also has the structure of this commutative ring, that is, the product is defined as

$$x \times y \;=\; \xi \cdot g \times \eta \cdot g \;=\; (\xi\eta)\cdot g \;=\; (\widetilde{\xi\eta})\cdot g.$$

If χ_g = irreducible, then X is even a field. Hence, in particular, X has a galois group. No one has ever given a dynamical interpretation of this galois group. In other words, there are obvious algebraic facts in the theory of dynamical systems which have never been examined from the dynamical point of view. For some related comments in the setting of topological semigroups, see DAY and WALLACE [1967].

6. TRANSFER FUNCTIONS

(6.0) PREAMBLE. There has been a vigorous tradition in engineering (especially in electrical engineering in the United States during 1940-1960) that seeks to phrase all results of the theory of linear constant dynamical systems in the language of the Laplace transform. Textbooks in this area often try to motivate their biased point of view by claiming that "the Laplace transform reduces the analytical problem of solving a differential equation to an algebraic problem". When directed to a mathematician, such claims are highly misleading because the mathematical ideas of the Laplace transform are never in fact used. The ideas which are actually used belong to classical complex function theory: properties of rational functions, the partial-fraction expansion, residue calculus, etc. More importantly, the word "algebraic" is used in engineering in an archaic sense and the actual (modern) algebraic content of engineering education and practice as related to linear systems is very meager. For example, the crucial concept of the transfer function is usually introduced via heuristic arguments based on linearity or "defined" purely formally as "the ratio of Laplace transforms of the output over the input". To do the job right, and to recognize the transfer function as a natural and purely algebraic gadget, requires a drastically new point of view, which is now at hand as the machinery set up in Sections 3-5. The essential idea of our present treatment was first published in KALMAN [1965b].

The first purpose of this section is to give an intrinsically algebraic definition of the transfer function associated with a discrete-time, constant, linear input/output map (see Definition (3.10)). Since the applications of transfer functions are standard, we shall not develop them in detail, but we do want to emphasize their role in relating the classical invariant factor theorem for polynomial matrices to the corresponding module theorem (4.34).

Consider an arbitrary $K[z]$-homomorphism $f\colon \Omega \to \Gamma$ (see lemma (g) following Theorem (4.2)). Then as a "mathematical object" f is equivalent to the set $\{f(e_j),\ i = 1, \ldots, m,\ e_j$ defined by (4.6)$\}$, since

$$(6.1) \qquad f(\omega) = \sum_{j=1}^{m} \omega_j \cdot f(e_j).$$

(The scalar product on the right is that in the $K[z]$-module Γ, as defined in Section 4.) By definition of Γ, each $f(e_j)$ is a formal power series in z^{-1} with vanishing first term. We shall try to represent these formal power series by ratios of polynomials (which we shall call <u>transfer functions</u>*) and then we can replace formula (6.1) by a certain specially defined product of a ratio of polynomials by a polynomial. Some algebraic sophistication will be needed to find the correct rules of calculations. These "rules" will consititute a rigorous (and simple) version of Heaviside's so-called "calculus". There are no conceptual complications of any sort. (However, we are dodging some difficulties by working solely in discrete-time.)

----

*This entrenched terminology is rather unenlightening in the present algebraic context.

Let $X_f = \Omega/\text{kernel } f$ be the state set of f regarded as a $K[z]$-module. We assume that X_f is a torsion module with nontrivial minimal polynomial ψ. Then, for each $j = 1, \ldots, m$ we have

$$(6.2) \qquad \psi \cdot f(e_j) = f(\psi \cdot e_j) = \eta([\psi \cdot e_j]) = \eta(\psi \cdot [e_j]) = 0.$$

By definition of the module structure on Γ, (6.2) means that the ordinary product of the power series $f(e_j)$ by the polynomial ψ is a (vector) polynomial. Hence (6.2) is equivalent to (notation: no dot = ordinary product)

$$(6.2') \qquad \psi f(e_j) = \theta_j \in K^p[z], \quad j = 1, \ldots, m.$$

Intuitively, we can solve this equation by writing $f(e_j) = \theta_j/\psi$. There are two ways of making this idea rigorous.

Method 1. Define

$$(6.3) \qquad f(e_j) = \theta_j/\psi$$

as the formal division of θ_j by ψ into ascending powers of z^{-1}. Check that the coefficient of z^0 is always 0. Verify by computation that the power series so obtained satisfies (6.2').

Method 2. Multiply both sides of (6.2') by z^{-m}. Write $\hat{\psi}(z^{-1}) = z^{-n}\psi(z)$ and $\hat{\theta}_j(z^{-1}) = z^{-n}\theta(z)$. Then $\hat{\psi} \in K[z^{-1}] \subset K[[z^{-1}]]$ and (6.2') becomes

$$(6.2'') \qquad \hat{\psi} f(e_j) = \hat{\theta}_j \in K^p[z^{-1}].$$

Moreover, the 0-th coefficient of $\hat{\psi}$ is 1 (because of the convention that the leading coefficient of ψ is 1), hence $\hat{\psi}$ is a unit in $K[[z^{-1}]]$ and therefore

$$(6.3') \qquad f(e_j) = \hat{\theta}(z^{-1})\hat{\psi}^{-1}(z^{-1}).$$

Note that (6.3) and (6.3') actually give slightly different definitions of $f(e_j)$, depending on whether we use a transfer function with respect to the variable z or z^{-1}. (Both notations have been used in the engineering literature.) For us the formalism of Method 1 is preferable. (The calculations of Method 1 can be reduced by Method 2 to the better-known calculations of the inverse in the ring $K[[z^{-1}]]$.)

Summarizing, we have the easy but fundamental result:

(6.4) EXISTENCE OF TRANSFER FUNCTIONS. <u>There is a bijective correspondence between</u> $K[z]$-<u>homomorphisms</u> $f: \Omega \to \Gamma$ <u>with minimal polynomial</u> ψ <u>and transfer function matrices of the type</u>

$$Z = [\theta_1/\psi, \ldots, \theta_m/\psi],$$

<u>where</u> $\theta_j \in K^p[z]$, $\deg \theta_j < \deg \psi$, <u>and</u> ψ <u>is the least common denominator of</u> Z.

In many contexts, it is preferable to deal with the Z_f corresponding to f rather than with f itself. Because the correspondence is bijective, it is clear that all objects induced by f are well-defined also for Z_f and conversely. Thus, for instance,

$$\dim Z_f \overset{\Delta}{=} \dim f \overset{\Delta}{=} \dim X_f;$$

$$\begin{aligned} \psi_Z &= \text{least common denominator of } Z, \\ &= \text{minimal polynomial of } f_Z. \end{aligned}$$

(6.5) REMARK. In view of Propositions (4.20-21), the natural realization of Z, namely $X_Z \triangleq X_{f_Z}$, is completely reachable as well as completely observable. Not having this fact available before 1960 has caused a great confusion. Questions such as those resolved by Theorem (5.13) tended to be attacked algorithmically, using special tricks amounting to elementary algebraic manipulations of elements of Z. Very few theoretical results could be conclusively established by this route until the conceptual foundations of the theory of reachability and observability were developed.

The preceding results may be restated as "rules" whereby the values of f may be computed using Z. We have in fact, $f(\omega) = Z\cdot\omega$, where

(6.6) $$Z\cdot\omega \triangleq (\widetilde{\psi Z\omega})/\psi,$$

= multiply the polynomial matrix ψZ consisting of the numerators of Z with ω, reduce to minimal-degree polynomials modulo ψ and then divide formally by ψ as in Method 1 above.

We can also compute the entire output of the system Σ_Z (that is, all output values following the application of the first nonzero input value) by the rule

(6.7) $$Z\omega \triangleq (\psi Z\omega)/\psi,$$

= same as above, but do not reduce modulo ψ.

In this second case, the output sequence will begin with a positive power of z. (The coefficients of the positive powers of z are thrown away in the definition of f (see (3.7)) and in the definition of the scalar product in Γ, in order to secure a simple formula for $X_f = \Omega/\text{kernel } f$.)

Many other applications of transfer functions may be found in KALMAN, FALB, and ARBIB [1969, Chapter 10, Section 10].

It is easy to show that the transfer function associated with the system $\Sigma_f = (F, G, H)$ is given by $Z_f = H(zI - F)^{-1}G$. (This is just the formal Laplace transform computed from the constant version of (1.12) by setting $z = d/dt$ or from (1.17) by setting $x(t + 1) = zx(t)$.) Probably the simplest way of computing Z is via the formula

$$6.8) \qquad (zI - F)^{-1} = \sum_{j=0}^{q-1} z^j \psi_F^{(q-i)}(F)(z), \qquad q = \deg \psi,$$

where ψ_F is the minimal polynomial of the matrix F and the superscript denotes the special polynomials defined in (5.5). The matrix identity (6.8) follows at once from the classical scalar identity [WEBER, 1898, §4]

$$\pi(z) - \pi(w) = (z - w) \sum_{j=1}^{q-1} z^j \pi^{(q-i)}(w), \qquad q = \deg \pi,$$

upon setting $w = F$, $\pi = \psi_F$, and invoking the Cayley-Hamilton theorem.

Much of classical linear system theory was concerned with computing Z_f. In the modern context, this problem "factors" into first solving the realization problem $f \to \Sigma_f$ and then applying formula (6.8). See Sections 8 and 9.

One of the mysterious features of Rule (6.6) (as contrasted with the conventional rule (6.7)) is the necessity of reducing modulo ψ. The simplest way of understanding the importance of this aspect of the problem is to show how to relate the module invariant factors occuring in the structure theorem (4.34) to the classical facts concerning the invariant factors of a polynomial matrix.

(6.9) INVARIANT FACTOR THEOREM FOR MATRICES. *Let* P *be a* $p \times m$ *matrix with elements in an arbitrary principal-ideal domain* R. *Then*

(6.10) $P = A\Pi B$,

where A *and* B *are* $p \times p$ *and* $m \times m$ *matrices (not necessarily unique) with elements in* R *and* $\det A$, $\det B$ *units in* R, *while*

(6.11) $\Pi = \operatorname{diag}(\lambda_1, \ldots, \lambda_q, 0, \ldots, 0)$ *with* $\lambda_i \in R$

is unique (up to units in R) *with* $\lambda_i | \lambda_{i+1}$, $i = 1, \ldots, q-1$, *and* $q = \operatorname{rank} P$. *The* λ_i *are called the invariant factors of* P.

As anyone would expect, there is a correspondence between the module structure theorem (4.34) and the matrix structure theorem (6.9) and, in particular, between the respective invariant factors $\psi_1, \ldots, \psi_r$ and $\lambda_1, \ldots, \lambda_q$. Let us sketch the standard proof of this fact following CURTIS and REINER [1962, §13.3] who also give a proof of (6.9).

PROOF OF (4.34). Consider the R-homomorphism from R^m onto M given by $\mu: e_i \mapsto g_i$, where the e_i are the standard basis elements of R^m (recall (4.6)) and the g_i generate M. Clearly, $M \approx R^s/N$, where $N = \text{kernel } \mu$. It can be proved that $N \approx R^\ell$ is a free submodule of R^m, with a basis of at most $\ell \leqq m$ elements. Write each basis element f_j of N as $\sum p_{ij} \cdot e_i$, $p_{ij} \in R$.

Apply (6.9) to the R-matrix P. Define $\hat{f}_j = \sum c_{ij}\cdot f_i$, $C = B^{-1}$, $\hat{e}_j = \sum a_{ij}\cdot e_i$. By (6.10-11), $\hat{f}_k = \lambda_i\cdot\hat{e}_i$. Hence

$$N = \lambda_1 R \oplus \dots \oplus \lambda_r R.$$

Then, by "direct sum",

$$M \approx R/\lambda_r R \oplus \dots \oplus R/\lambda_1 R \oplus R^{m-\ell}, \quad i = 1, \dots, r.$$

That is, (4.34) holds with $\psi_i = \lambda_i$ and $r = \text{rank } P = \ell$. □

By the same type of calculations, we can prove also

(6.12) THEOREM. *Let* $\lambda_1, \dots, \lambda_q$ *be the invariant factors of* ψZ *given by* (6.9), *and let* $(\lambda_i, \psi) = \theta_i$, $i = 1, \dots, q$. *Then the invariant factors of* X_Z *are*

$$\begin{aligned} \psi_1 &= \psi, \\ \psi_2 &= \psi/\theta_2, \\ &\ \ \vdots \\ \psi_r &= \psi/\theta_r, \end{aligned}$$

where r *is the smallest integer such that* $\psi \mid \lambda_i$ *for* $i = r + 1, \dots, q = \text{rank } \psi Z$.

PROOF. Consider the $K[z]$-epimorphism $\mu: \Omega \to X_Z: \omega \to [\omega]_Z$. Clearly, $\omega \in [0]_Z = \text{kernel } \mu$ iff $Z\cdot\omega = 0$ (see (6.6)). Equivalently, $(\psi Z)\omega = 0 \pmod{\psi}$. Using the representation whose existence is claimed by (6.9), write $\psi Z = C\Lambda D$ (C, Λ, D = matrices over $K[z]$.) Define $W = D^{-1}\Psi$, where

$$\Psi = \text{diag}\,(\psi_1, \psi_2, \ldots, \psi_r, 1, \ldots, 1).$$

Then $\Lambda\Psi = 0$, $(\psi Z)W = 0$, and W has clearly maximal rank among $K[z]$-matrices with this property. So the columns of the matrix W constitute a basis for kernel μ. The rest follows easily, as in the proof of (4.34). □

(6.13) REMARK. The preceding proof remains correct, without any modification, if the representation $\psi Z = C\Lambda D$, det C, det D = units is taken in the ring $K[z]/\psi K[z]$, rather than in $K[z]$. The former representation follows trivially from the latter but may be easier to compute.

(6.14) REMARK. Theorem (6.12) shows how to compute the invariant factors of X_Z from those of ψZ. We must define the invariant factors of Z to be the same as those of X_Z (because of the bijective correspondence $Z \leftrightarrow X_Z$). Consistency with (6.12) demands that we write

$$(6.15) \qquad \lambda_i/\psi = (\lambda_i/\theta_i)/(\psi/\theta_i), \qquad \theta_i = (\lambda_i, \omega),$$

where $/$ is defined as in (6.3). In other words, the ψ_i are the denominators of the scalar transfer function λ_i/ψ after cancellation of all common factors.

Theorems (4.34) and (6.12) do not fully reveal the significance of invariant factors in dynamical systems. Nor is it convenient to deduce all properties of matrix-invariant factors from the representation theorem (6.9). It is interesting that the sharpened results we present below are much in the spirit of the original work of WEIERSTRASS, H. J. S. SMITH, KRONECKER, FROBENIUS, and HENSEL, as summarized in the well-known monograph of MUTH [1899].

(6.16) DEFINITION. *Let* A, B *rectangular matrices over a unique factorization domain* R. $A|B$ (*read:* A *divides* B) *iff there are matrices* V, W (*over* R, *of appropriate sizes*) *such that* $B = VAW$.

This is of course just the usual definition of "divide" in a ring, specialized to the noncommutative ring of matrices.

The following result [MUTH 1899, Theorems IIIa-b, p. 52] shows that in case of principal-ideal domains the correspondence between matrices and their invariant factors preserves the divide relation (is "functorial" with respect to "divide"):

(6.17) THEOREM. *Let* R *be a principal-ideal domain. Then* $A|B$ *if and only if* $\lambda_i(A) | \lambda_i(B)$ *for all* i.

PROOF. *Sufficiency.* Write the representation (6.10) as

$$A = V_1 \Lambda_1 W_1, \qquad B = V_2 \Lambda_2 W_2.$$

By hypothesis, there is a Λ_3 (diagonal) such that $\Lambda_1 \Lambda_3 = \Lambda_2$. Hence

$$\begin{aligned} B &= V_2 \Lambda_1 \Lambda_3 W_2, \\ &= V_2 V_1^{-1} A V_1 \Lambda_1 W_1 W_1^{-1} \Lambda_3 W_2, \\ &= (V_2 V_1^{-1}) A (W_1^{-1} \Lambda_3 W_2). \end{aligned}$$

Necessity. This is just the following

(6.18) LEMMA. For an arbitrary unique-factorization domain R, $A|B$ implies $\lambda_i(A)|\lambda_i(B)$.

PROOF. By elementary determinant manipulations, as in MUTH [1899, Theorem II, p. 16-17]. □

This completes the proof of Theorem (6.17) □

(6.19) REMARK. Since (6.9) does not apply (why?) to unique factorization domains, for purposes of using Lemma (6.18) we need WEIERSTRASS's definition of invariant factors: if $\Delta_j(A)$ = greatest common factor of all $j \times j$ minors of a matrix A, with $\Delta_0(A) = 1$, then $\lambda_i(A) = \Delta_i(A)/\Delta_{i-1}(A)$. Of course, this definition can be shown to be equivalent (over principal-ideal domains) to that implied by (6.9).

In analogy with Definition (6.16), let us agree (note inversion!) on

(6.20) DEFINITION. Let Z_1, Z_2 be transfer-function matrices $Z_1|Z_2$ (read: Z_1 divides Z_2) iff there are matrices V, W over $K[z]$ such that $Z_1 = VZ_2W$. (Note that $Z_1|Z_2$ implies at once: $\psi_{Z_1}|\psi_{Z_2}$.)

(6.21) THEOREM. $Z_1|Z_2$ if and only if $\psi_i(Z_1)|\psi_i(Z_2)$ for all i.

PROOF. This is the natural counterpart of Theorem (6.16), and follows from it by a simple calculation using the definition of $\psi_i(Z)$ given by (6.15). □

(6.22) DEFINITION. $\Sigma_1 | \Sigma_2$ (read: Σ_1 can be simulated by Σ_2) iff $X_{\Sigma_1} | X_{\Sigma_2}$, that is, iff X_{Σ_1} is isomorphic to a submodule of X_{Σ_2} [or isomorphic to a quotient module of X_{Σ_2}].

This definition is also functorially related to the definition of "divide" over a principal ideal domain R because of the following standard result:

(6.23) THEOREM. Let R be a principal-ideal domain and X, Y R-modules. Then Y is (isomorphic) to a submodule or quotient module of X if and only if

$$\psi_i(Y) | \psi_i(X), \qquad i = 1, \ldots, r(Y) \leqq r(X).$$

PROOF. Sufficiency. Take both X and Y in canonical form (4.34), with $x_1, \ldots, x_{r(X)}$ generating the cyclic pieces of X, and $y_1, \ldots, y_{r(X)}$ (with $y_i = 0$ if $i > r(Y)$) those of Y. The assignment $y_i \mapsto (\psi_i(X)/\psi_i(Y))\cdot x_i$ defines a monomorphism $Y \to X$, that is, exhibits Y as (isomorphic to) a submodule of X. Similarly, the assignment $x_i \mapsto y_i$ defines an epimorphism $X \to Y$ exhibiting Y as (isomorphic to) a quotient module of X.

Necessity (following BOURBAKI [Algèbre, Chapter 7 (2^e ed.), Section 4, Exercise 8]). Let Y be a submodule of X. By (4.34), $X \approx L/N$ where L, N are free R-modules. By a classical isomorphism theorem, Y is isomorphic to a quotient module M/N, where $L \supset M \supset N$ and M is free (since submodules of a free module are free).

From the last relation, $r(Y) \leqq r(X)$. Now observe, again using (4.34) that, for any R-module X and any $\pi \in R$,

$$r(\pi X) < k \implies \psi_k(X) | \pi$$

and therefore

$$R\psi_k(X) = \text{ideal generated by } \{\pi : r(\pi X) < k\}.$$

Since πY is a submodule of πX for all $\pi \in R$, it follows that $R\psi_k(X) \supset R\psi_k(Y)$, and the proof is complete for the case when Y is a submodule of X. The proof of the other case is similar. □

(6.24) COROLLARY. $\psi_i(Z_\Sigma) | \psi_i(\Sigma)$, $i = 1, \ldots, r(Z_\Sigma)$.

PROOF. Immediate from the fact that X_{Z_Σ} is a submodule of Σ (see Section 7). □

Now we can summarize main results of this section as the

(6.25) PRIME DECOMPOSITION THEOREM FOR LINEAR DYNAMICAL SYSTEMS. The following conditions are equivalent:

(i) Z_1 divides Z_2.

(ii) $\psi_i(Z_1)$ divides $\psi_i(Z_2)$ for all i.

(iii) Σ_{Z_1} can be simulated by Σ_{Z_2}.

PROOF. This follows by combining Theorem (6.21) with Theorem (6.23), since $\psi_i(Z) = \psi_i(\Sigma_Z)$ by definition. □

(6.26) INTERPRETATION. The definition of $Z_1 | Z_2$ means, in system-theoretic terms, that the inputs and outputs of the machine whose transfer function is Z_2 are to be "recoded": the original input ω_2 is replaced by an input $\omega_2 = B(z)\omega_1$ and the output γ_2 is replaced by an output $\gamma_1 = A(z)\gamma_2$; with these "coding" operations, Σ_2 will act like a machine with transfer function Z_1. In view of the definition of a transfer function, the equation $Z_1 = AZ_2B$ is always satisfied whenever A, B are replaced by $\tilde{A}$, $\tilde{B}$ (reduced modulo ψ_{Z_2}). This means that the coding operations can be carried out physically given a delay of $d = \deg \psi_{Z_2}$ units of time (or more). No feedback is involved in coding, it is merely necessary to store the d last elements of the input and output sequences. Hence, in view of Theorem (6.25) and Corollary (6.24), we can say that it is possible to alter the dynamical behavior of a system Σ_2 arbitrarily by external coding involving delay but not feedback if and only if the invariant factors of the desired external behavior (Z_1) are divisors of invariant factors of the external behavior (Z_{Σ_2}) of the given system. The invariant factors may be called the PRIMES of linear systems: they represent the atoms of system behavior which cannot be simulated from smaller units using arbitrary but feedback-free coding. In fact, there is a close (bot not isomorphic) relationship between the Krohn-Rhodes primes of automata theory (see KALMAN, FALB, and ARBIB [1969, Chapters 7-9]) and ours. A full treatment of this part of linear system theory will be published elsewhere.

7. ABSTRACT THEORY OF REALIZATIONS

The purpose of this short section is to review and expand those portions of the previous discussion which are relevant to the detailed theory of realizations to be presented in Sections 8 and 9. The same issues are examined (from a different point of view) also in KALMAN, FALB, and ARBIB [1969].

Let $f: \Omega \to \Gamma$ be a fixed input/output map. Let us recall the construction of X_f, as a set and as carrying a $K[z]$-module structure (Sections 3 and 4). It is clear that (i) $f = \iota_f \circ \mu_f$, where

$$\mu_f: \Omega \to X_f: \omega \mapsto [\omega],$$
$$\iota_f: X_f \to \Gamma: [\omega]_f \mapsto f(\omega)$$

are $K[z]$-homomorphisms, and (ii) μ_f = epimorphism while ι_f = monomorphism. We have also seen that

$$(7.1) \quad \begin{cases} \mu_f = \text{epimorphism} \iff X_f \text{ is completely reachable;} \\ \\ \iota_f = \text{monomorphism} \iff X_f \text{ is completely observable.} \end{cases}$$

These facts set up a "functor" between system-theoretic notions and algebra which characterize X_f uniquely. Consequently, it is desirable to replace also our system-theoretic definition of a realization (3.12) by a purely algebraic one:

(7.2) DEFINITION. *A realization of a* $K[z]$-*homomorphism* $f: \Omega \to \Gamma$ *is any factorization* f *that is, any commutative diagram*

$$\begin{tikzcd}
\Omega \arrow[rr, "f"] \arrow[dr, "\mu"'] & & \Gamma \\
& X \arrow[ur, "\iota"'] &
\end{tikzcd}$$

of $K[z]$-homomorphisms. The $K[z]$-module X is called the state module of the realization. A realization is canonical iff it is completely reachable and completely observable, that is, μ is surjective and ι is injective.

A realization always exists because we can take $X = \Omega$, $\mu = 1_{\Omega}$, $\iota = f$ (or $X = \Gamma$, $\mu = f$, $\iota = 1_{\Gamma}$).

(7.3) REMARK. It is clear that a realization in the sense of (3.12) can always be obtained from a realization given by (7.2). In fact, define $\Sigma = (F, G, H)$ by

$$F\colon X \to X\colon x \mapsto z \cdot x,$$

$G = \mu$ restricted to the submodule $\{\omega\colon |\omega| = 1\}$.

$H = \iota$ followed by the projection $\gamma \mapsto \gamma(1)$.

It is easily verified that these rules will define a system with f, x = f. Given any such Σ, it is also clear that the rules

$$X = X_{\Sigma},$$

$$\mu\colon \omega \mapsto \sum_{t \leqq 0} F_{\Sigma}^{-t} G_{\Sigma}\omega(t),$$

$$\nu\colon x \mapsto (H_{\Sigma}x, H_{\Sigma}F_{\Sigma}x, \dots)$$

define a factorization of f. Hence the correspondence between (3.12) and (7.2) is bijective.

The quickest way to exploit the algebraic consequences of our definition (7.2) is via the following arrow-theoretic fact:

(7.4) ZEIGER FILL-IN LEMMA. *Let* A, B, C, D *be sets and* α, β, γ, *and* δ *set maps for which the following diagram commutes:*

$$\begin{tikzcd} A \arrow[r, "\alpha"] \arrow[d, "\gamma"'] & B \arrow[d, "\beta"] \arrow[dl, dashed] \\ C \arrow[r, "\delta"'] & D \end{tikzcd}$$

If α *is surjective and* δ *is injective, there exists a unique set map* φ *corresponding to the dashed arrow which preserves commutativity*.

This follows by straightforward "diagram-chasing", which proves at the same time the

(7.5) COROLLARY. *The claim of the lemma remains valid if "sets" are replaced by "R-modules" and "set maps" by "R-homomorphisms"*.

Applying the module version of the lemma twice, we get

(7.6) PROPOSITION. *Consider any two canonical realizations of a fixed* f: *the corresponding state-sets are isomorphic as* K[z]*-modules*.

Since every K[z]-module is automatically also a K-vector space, (7.6) shows that the two state sets are K-isomorphic, that is, have the same dimension as vector spaces. The fact that they are also K[z]-isomorphic implies, via Theorem (4.34), that they have the same invariant factors. We have already employed the convention that (in view of the bijection between f and Σ_f), the invariant factors of f and X_f are to be identified. In view of (7.6), this is now a general fact, not dependent on the special construction used to get X_f. We can therefore restate (7.6) as the

(7.7) ISOMORPHISM THEOREM FOR CANONICAL REALIZATIONS. *Any two canonical realizations of a fixed* f *have isomorphic state modules. The state module of a canonical realization is uniquely characterized (up to isomorphism) by its invariant factors, which may be also viewed as those of* f.

A simple exercise proves also

(7.8) PROPOSITION. *If* X *is the state module of a canonical realization* f, *then* $\dim X$ *(as a vector space) is minimum in the class of all realizations of* f.

This result has been used in some of the literature to justify the terminology "minimal realization" as equivalent to "canonical realization". We shall see in Section 9 that the two notions are not always equivalent; we prefer to view (7.2) as the basic definition and (7.8) as a derived fact.

(7.9) REMARK. Theorem (7.7) constitutes a proof of the previously claimed (4.24). To be more explicit: if $\Sigma = (F, G, H)$ and $\hat{\Sigma} = (\hat{F}, \hat{G}, \hat{H})$ are two triples of matrices defining canonical realizations of the same f, then (7.7) implies the existence of a vector-space isomorphism $A: X \rightarrow \hat{X}$ such that

$$\begin{aligned} \hat{F} &= AFA^{-1}, \\ \hat{G} &= AG, \\ \hat{H} &= HA^{-1}. \end{aligned} \tag{7.10}$$

If we identify X and $\hat{X}$ then A is simply a basis change and it follows that *the class of all matrix triples which are canonical realizations of a fixed* f *is isomorphic with the general linear group over* X.

The actual computation of a canonical realization, that is, of the abstract Nerode equivalence classes $[\omega]_f$, require a considerable amount of applied-mathematical machinery, which will be developed in the next section. The critical hypthesis is the existence of a factorization of f such that $\dim X < \infty$. (this is sometimes expressed by saying that f has *finite rank*.) Given any such realization, it is possible to obtain a canonical one by a process of reduction. More precisely, we have

(7.11) THEOREM. *Every realization of* f *with state module* X *contains a subquotient (a quotient of a submodule, or equivalently, a submodule of a quotient)* X_* *of* X *which is the state-module of a canonical realization of* f.

PROOF. The reachable states X_r = image μ are a submodule of X and so are the unobservable states X_o = kernel ι. Hence $X_* \approx X_r/X_r \cap X_o$ is a subquotient of X. It follows immediately that X_* is a canonical state-module for f. [The proof may be visualized via the following commutative diagram, where the j's and p's are canonical injections and projections.] □

$$\begin{tikzcd}
\Omega \arrow[rr, "f"] \arrow[d, "\mu_*"'] \arrow[dr, "\mu"] & & \Gamma \\
X_r \arrow[r, "j_1"'] \arrow[dr, "p_2"'] & X \arrow[r, "p_1"'] \arrow[ur, "\iota"] & X/X_{\bar{o}} \arrow[u, "\iota_*"'] \\
& X_r/X_r \cap X_{\bar{o}} \arrow[ur, "j_2"'] &
\end{tikzcd}$$

(7.12) REMARK. Since any subquotient of X is isomorphic to a submodule (or a quotient module) of X, it follows from Theorem (6.23) that X can be state-state module of a realization only if $\psi_i(f) | \psi_i(X)$ for all i (recall also Corollary (6.24)). This condition, however, is not enough since the ψ_i are invariants of module isomorphisms and not isomorphisms of the commutative diagram (7.2).

The preceding discussion should be kept in mind to gain an overview of the algorithms to be developed in the next sections.

8. CONSTRUCTION OF REALIZATIONS

Now we shall develop and generalize the basic algorithm, originally due to B. L. Ho (see HO and KALMAN [1966]), for computing a canonical realization $\Sigma = (F, G, H)$ of a given input/output map f. Most of the discussion will be in the language of matrix algebra.

Notations. Here and in Section 9 boldface capital letters* will denote block matrices or sequences of matrices; finite block matrices will be denoted by small Greek subscripts on boldface capitals; the elements of such matrices will be denoted by ordinary capitals. This is intended to make the practical aspects of the computations self-evident; no further explanations will be made.

Let $f: \Omega \rightarrow \Gamma$ be a given, fixed $K[z]$-homomorphism. Using only the K-linearity of f we have that

$$(8.1) \qquad f(\omega)(1) \quad = \sum_{t \leqq 0} A_{-t+1}\omega(t),$$

where the A_k $(k > 0)$ are $p \times m$ matrices over the fixed field K. We denote the totality of these matrices by

$$\underline{\underline{A}}(f) \quad = \quad (A_1, A_2, \ldots).$$

Then it is clear that *the specification of a* $K[z]$-*homomorphism* f *is equivalent to the specification of its matrix sequence* $\underline{\underline{A}}(f)$. Moreover, if Σ realizes f (8.1) can be written explicitly as

$$(8.2) \qquad f(\omega)(1) \quad = \sum_{t \leqq 0} HF^{-t}G\omega(t).$$

----

*Note to Printer: Indicated by *double* underline.

Comparing (8.1) and (8.2) we can translate (3.12) into an equivalent matrix-language

(8.3) DEFINITION. *A dynamical system* $\Sigma = (F, G, H)$ *realizes a (matrix) infinite sequence* $\underline{\underline{A}}$ *iff the relation*

$$A_{k+1} = HF^kG, \quad k = 0, 1, 2, \dots$$

is satisfied.

Let us now try to obtain also a matrix criterion for an infinite sequence $\underline{\underline{A}}$ to have a finite-dimensional realization. The simplest way to do that is to first write down a matrix representation for the map $f: \Omega \to \Gamma$. So let

$$\underline{\underline{H}}(\underline{\underline{A}}) = \begin{bmatrix} A_1 & A_2 & A_3 & \dots \\ A_2 & A_3 & A_4 & \dots \\ A_3 & A_4 & A_5 & \dots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{bmatrix},$$

and verify that $\underline{\underline{H}}(\underline{\underline{A}}(f))$ represents f when $\omega \in \Omega$ is viewed as an ∞ column vector with elements $(\omega_1(0), \dots, \omega_m(0), \omega_1(1), \dots)$. Classically, $\underline{\underline{H}}(\underline{\underline{A}})$ is known as the (infinite) *Hankel matrix* associated with $\underline{\underline{A}}$. We denote by $\underline{\underline{H}}_{\mu,\nu}$ the $\mu \times \nu$ block submatrix of $\underline{\underline{H}}$ appearing in the upper left-hand corner of $\underline{\underline{H}}$.

(8.4) PROPOSITION. *Let* Σ *be any realization of* $\underline{\underline{A}}$. *Then*

$$\text{rank } \underline{\underline{H}}_{\mu,\nu}(\underline{\underline{A}}) \leq \dim \Sigma \quad \textit{for all} \quad \mu, \nu \geq 1.$$

(8.5) COROLLARY. An infinite sequence $\underline{\underline{A}}$ has a finite-dimensional realization only if rank $\underline{\underline{H}}_{\mu,\nu}(\underline{\underline{A}})$ is constant for all μ, ν sufficiently large.

PROOF. If $\dim \Sigma = \infty$, the claim of the proposition is vacuous (although formally correct!). Assume therefore that $\dim \Sigma < \infty$ and define from Σ the finite block matrices

$$\underline{\underline{R}}_\nu = [G, FG, \ldots, F^{\nu-1}G]$$

and

$$\underline{\underline{O}}_\mu = [H', H'F', \ldots, H'(F')^{\mu-1}].$$

Then

$$\underline{\underline{O}}'_\mu \underline{\underline{R}}_\nu = \underline{\underline{H}}_{\mu,\nu}(\underline{\underline{A}})$$

by the definition (8.3) of a realization. It is clear that rank $\underline{\underline{R}}_\nu$ and rank $\underline{\underline{O}}_\mu$ are at most $n = \dim \Sigma$. Thus our claim is reduced to the standard matrix fact

$$\text{rank } (AB) \leqq \min \{\text{rank } A, \text{ rank } B\}. \qquad \square$$

Our next objective is the proof of the converse of the corollary. This can be done in several ways. The original proof is due to HO and KALMAN [1966]; similar results were obtained independently and concurrently by YOULA and TISSI [1966] as well as by SILVERMAN [1966]. Two different proofs are analyzed and compared in KALMAN, FALB, and ARBIB [1969, Chapter 10, Section 11]. All proofs depend on certain finiteness arguments. We shall give here a variant of the proof developed in HO and KALMAN [1969].

(8.6) DEFINITION. *The infinite Hankel matrix* $\underline{\underline{H}}$ *associated with the sequence* $\underline{\underline{A}}$ *has finite length* $\lambda = (\lambda', \lambda'')$ *iff one of the following two equivalent conditions holds:*

$$\lambda' = \min \{\ell' : \operatorname{rank} \underline{\underline{H}}_{\ell',\nu} = \operatorname{rank} \underline{\underline{H}}_{\ell'+\kappa,\nu} \text{ for all } \kappa, \nu = 1, 2, \ldots \} < \infty$$

or

$$\lambda'' = \{\min \ \ell'' : \operatorname{rank} \underline{\underline{H}}_{\mu,\ell''} = \operatorname{rank} \underline{\underline{H}}_{\mu,\ell''+\kappa} \text{ for all } \kappa, \mu = 1, 2, \ldots \} < \infty.$$

λ' *is the row length of* $\underline{\underline{H}}$ *and* λ'' *is the column length of* $\underline{\underline{H}}$.

The equivalence of the two conditions is immediate from the equality of the row rank and column rank of a finite matrix. The proof of the following result (not needed in the sequel) is left for the reader as an exercise in familiarizing himself with the special pattern of the elements of a Hankel matrix:

(8.7) PROPOSITION. *For any* $\underline{\underline{H}}$, *the following inequalities are either both true* [$\underline{\underline{H}}$ *has finite length*] *or both false* [*otherwise*]:

$$\lambda' \leqq \operatorname{rank} \underline{\underline{H}}_{m\lambda'',\lambda''} \leqq m\lambda'',$$
$$\lambda'' \leqq \operatorname{rank} \underline{\underline{H}}_{\lambda',p\lambda'} \leqq p\lambda',$$

The most direct consequence of the finiteness condition given by (8.6) is the existence of a finite-dimensional representation $\underline{\underline{S}}$ and $\underline{\underline{Z}}$ of the shift operator σ_A acting on a sequence $\underline{\underline{A}}$. The "operand" will be the Hankel matrix associated with a given $\underline{\underline{A}}$. As we shall see soon, this representation of the shift operator induces a rule for computing the matrix F of a realization of $\mathbf{A}$. This is exactly what we would expect: module theory tells us that, loosely speaking,

$$\mathbf{S}, \mathbf{Z} \approx \sigma_A \approx z \approx F.$$

(8.7) DEFINITION. *The shift operator* σ_A *on an infinite sequence* $\mathbf{A}$ *is given by*

$$\sigma_A^k: (A_1, A_2, \dots) \mapsto (A_{1+k}, A_{2+k}, \dots);$$

the corresponding shift operator on Hankel matrices is then

$$\sigma_H^k: \mathbf{H}(\mathbf{A}) \mapsto \mathbf{H}(\sigma_A^k \mathbf{A}).$$

(Of course, σ_H is well-defined also on submatrices of a Hankel matrix.)

(8.8) MAIN LEMMA. *A Hankel matrix* $\mathbf{H}$ *associated with an infinite sequence* $\mathbf{A}$ *has finite length if and only if the shift operator* σ_H *has finite-dimensional left and right matrix representations*. *Precisely*: $\mathbf{H}$ *has finite length* $\lambda = (\lambda', \lambda'')$ *if and only if there exist* $\ell' \times \ell'$ *and* $\ell'' \times \ell''$ *block matrices* $\mathbf{S}$ *and* $\mathbf{Z}$ *such that*

$$\begin{aligned} (8.9) \qquad \sigma_H^k \mathbf{H}_{\ell',\ell''}(\mathbf{A}) &= \mathbf{S}\,\mathbf{H}_{\ell',\ell''}(\mathbf{A}), \\ &= \mathbf{H}_{\ell',\ell''}(\mathbf{A})\mathbf{Z}^k, \end{aligned}$$

and furthermore the minimum size of these matrices satisfying (8.9) *is* $\lambda' \times \lambda'$ *and* $\lambda'' \times \lambda''$.

PROOF. *Sufficiency*. Take any $\ell'' \times \ell''$ block matrix $\mathbf{Z}$ which satisfies (8.9). Compute the last column of $\mathbf{H}_{\mu,\ell''}\mathbf{Z}$:

$$(8.10) \qquad A_{j+\ell''+1} = A_{j+1}Z_{1\ell''} + A_{j+2}Z_{2\ell''} + \dots + A_{j+\ell''}Z_{\ell''\ell''}$$

for all $j = 0, 1, \dots$ (where $Z_{\mu\nu}$ is the $(\mu, \nu)^{\text{th}}$ element block of $\underline{\underline{Z}}$). Relation (8.10) proves that

$$\operatorname{rank} \underline{\underline{H}}_{\kappa+1,\ell''} = \operatorname{rank} \underline{\underline{H}}_{\kappa+1,\ell''} \quad \text{for all} \quad \kappa = 0, 1, \dots ;$$

the general case follows by repetition of the same argument. Hence the existence of the claimed $\underline{\underline{Z}}$ implies that the column length λ'' of $\underline{\underline{H}}$ cannot exceed the size of $\underline{\underline{Z}}$. If actually λ'' is smaller than the size of the smallest $\underline{\underline{Z}}$ which works in (8.9), we get a contradiction from the necessity part of the proof. The claims concerning $\underline{\underline{S}}$ are proved by a strictly dual argument.

<u>Necessity</u>. By the definition of λ'', each column of the $(\lambda'' + 1)^{\text{th}}$ block column of $\underline{\underline{H}}_{\mu,\lambda''+1}$ is linearly dependent on the columns of the preceding block columns of $\underline{\underline{H}}_{\mu,\lambda''+1}$; moreover, this property is true for all integers μ, no matter how large. So there exist $m \times m$ matrices $Z_1, \dots, Z_{\lambda''}$ such that the relation

$$(8.11) \qquad A_{j+1}Z_{\lambda''} + A_{j+2}Z_{\lambda''+1} + \dots + A_{j+\lambda''}Z_1 = A_{j+1+\lambda''}$$

holds identically for all $j = 0, 1, \dots$. Now define $\underline{\underline{Z}}$ to be an $\lambda'' \times \lambda''$ block companion matrix of $m \times m$ block made up from the Z_i just defined:

$$Z = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & Z_{\lambda''} \\ I & 0 & 0 & \cdots & 0 & Z_{\lambda''-1} \\ 0 & I & 0 & \cdots & 0 & Z_{\lambda''-2} \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 0 & Z_2 \\ 0 & 0 & 0 & \cdots & I & Z_1 \end{bmatrix} .$$

The verification of (8.9) is immediate, using (8.11). The existence of $\lambda' \times \lambda'$ block matrix $\underline{\underline{S}}$ verifying (8.9) follows by a strictly dual argument. □

Now we have enough material on hand to prove the strong version of Corollary (8.5):

(8.12) THEOREM. *An infinite sequence* $\underline{\underline{A}}$ *has a finite-dimensional realization of dimension* n *if and only if the associated Hankel matrix* $\underline{\underline{H}}$ *has finite length* $\lambda = (\lambda', \lambda'')$.

PROOF. *Sufficiency*. Let $\underline{\underline{E}}_{\lambda'',1}$ be a $\lambda'' \times 1$ block column matrix whose first block element is an $m \times m$ unit matrix and the other blocks are $m \times m$ zero matrices. Using (8.9) with $\ell'' = \lambda''$, define

$$(8.13) \qquad \Sigma = \begin{cases} F = \underline{\underline{Z}}, \\ G = \underline{\underline{E}}_{\lambda'',1}, \\ H = \underline{\underline{H}}_{1,\lambda''}. \end{cases}$$

Then, for all $k \geqq 0$, compute

$$\begin{aligned} HF^kG &= \underline{\underline{H}}_{1,\lambda''}\underline{\underline{Z}}^k\underline{\underline{E}}_{\lambda'',1} \\ &= \sigma^k_H\underline{\underline{H}}_{1,\lambda''}\underline{\underline{E}}_{\lambda'',1}; \end{aligned}$$

the second step uses (8.9). By definition of σ_A and $\underline{\underline{E}}$, the last matrix is just the $(1, 1)^{th}$ element of $\underline{\underline{H}}(\sigma^k_A(\underline{\underline{A}}))$, namely A_{1+k}. Hence the given Σ is a realization of $\underline{\underline{A}}$.

<u>Necessity</u>. This is immediate from Corollary (8.5). □

Now we want to attack the problem of finding a canonical realization of $\underline{\underline{A}}$, since the realization given by (8.13) is usually very far from canonical. Our succeeding considerations here and in Section 9 are made more transparent if we digress for a moment to establish another consequence of (8.8).

By outrageous abuse of language, we shall say that $\underline{\underline{A}}$ has <u>finite length</u> iff $\underline{\underline{H}}(\underline{\underline{A}})$ has finite length. We note

(8.14) DEFINITION. <u>An infinite sequence</u> $\underline{\underline{B}}$ <u>is an extension of order</u> N <u>of (the initial part of) an infinite sequence</u> $\underline{\underline{A}}$ <u>iff</u> $A_k = B_k$ <u>for</u> $k = 1, \ldots, N$.

(8.15) THEOREM. <u>No infinite sequence of finite length</u> (λ', λ'') <u>has distinct length-preserving extensions of any order</u> $N \geqq \lambda' + \lambda''$.

PROOF. Suppose $\underline{\underline{B}}$ is a length-preserving extension of order N of $\underline{\underline{A}}$, the length of both sequences being (λ', λ''), with $N \geqq \lambda' + \lambda''$. By (8.8), both sequences satisfy relation (8.9), with suitable $\underline{\underline{S}}_A$ and $\underline{\underline{Z}}_B$.

The sequence $\mathbf{A}$ is uniquely determined by $\mathbf{S}_A$ acting on $\mathbf{H}_{\lambda',\lambda''}(\mathbf{A})$ from the left and the sequence $\mathbf{B}$ is uniquely determined by $\mathbf{Z}_B$ acting on the matrix $\mathbf{H}_{\lambda',\lambda''}(\mathbf{B})$ from the right. The two matrices are equal by hypothesis on N. Moreover,

$$\mathbf{S}_A\mathbf{H}_{\lambda',\lambda''}(\mathbf{A}) = \sigma_A\mathbf{H}_{\lambda',\lambda''}(\mathbf{A})$$

and

$$\mathbf{H}_{\lambda',\lambda''}(\mathbf{B})\mathbf{Z}_B = \sigma_B\mathbf{H}_{\lambda',\lambda''}(\mathbf{B})$$

are *also* equal, since the matrices on the right-hand side depend only on the 2nd, ..., N-th member of each sequence. Using only this fact and the associativity of the matrix product

$$\begin{aligned}\mathbf{H}_{\lambda',\lambda''}\mathbf{Z}_B^k &= \mathbf{H}_{\lambda',\lambda''}\mathbf{Z}_B\mathbf{Z}_B^{k-1},\\ &= \mathbf{S}_A\mathbf{H}_{\lambda',\lambda''}\mathbf{Z}_B^{k-1},\\ &\quad\cdot\\ &\quad\cdot\\ &\quad\cdot\\ &= \mathbf{S}_A^{k-1}\mathbf{H}_{\lambda',\lambda''}\mathbf{Z}_B,\\ &= \mathbf{S}_A^k\mathbf{H}_{\lambda',\lambda''}.\end{aligned}$$

So $\mathbf{A} = \mathbf{B}$. □

Now we can hope for a realization algorithm which uses only the first $\lambda' + \lambda''$ terms of a sequence of finite length. In fact, we have

(8.16) B. L. HO's REALIZATION ALGORITHM. *Consider any infinite sequence* $\mathbf{A}$ *of finite length with associated Hankel matrix* $\mathbf{H}$. *The following steps will lead to a canonical realization of* $\mathbf{A}$:

(i) *Determine* λ', λ''.

(ii) *Compute* $n = \operatorname{rank} \underline{\underline{H}}_{\lambda',\lambda''}$; *in doing so, determine nonsingular* $p\lambda' \times p\lambda'$ *and* $m\lambda'' \times m\lambda''$ *matrices* P, Q *such that*

$$PH_{\lambda',\lambda''}Q = \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}. \tag{8.17}$$

(iii) *Compute*

$$\left\{\begin{aligned} F &= R_nP[\sigma_H\underline{\underline{H}}_{\lambda',\lambda''}]QC^n, \\ G &= R_nP\underline{\underline{H}}_{\lambda',\lambda''}C^m, \\ H &= R_p\underline{\underline{H}}_{\lambda',\lambda''}QC^n, \end{aligned}\right. \tag{8.18}$$

where R_p, C^m *are idempotent "editing" matrices corresponding to the operations "retain only the first* p *rows" and "retain only the first* m *columns".*

We claim the

(8.19) REALIZATION THEOREM FOR INFINITE SEQUENCES. *For any infinite sequence* $\underline{\underline{A}}$ *whose associated Hankel matrix* $\underline{\underline{H}}$ *has finite length* (λ', λ''), *B. L. Ho's formulas* (8.17-18) *yield a canonical realization.*

PROOF. If Σ defined by (8.17-18) is a realization of $\underline{\underline{A}}$, then it is certainly canonical: by (8.4) Σ has minimal dimension in the class of all realizations of $\underline{\underline{A}}$ and so it is canonical by (7.8).

The required verification is interesting. First, drop all subscripts. Observe that $\underline{\underline{H}}^{\#} = QCRP$ is a *pseudo-inverse* of $\underline{\underline{H}}$, that is, $\underline{\underline{H}}\underline{\underline{H}}^{\#}\underline{\underline{H}} = \underline{\underline{H}}$. Then, by definition of F, G, H, and $\underline{\underline{H}}^{\#}$,

$$\begin{aligned} HF^{k}G &= (R\underline{\underline{H}}QC)(RP[\sigma_{H}\underline{\underline{H}}]QC)^{k}(RP\underline{\underline{H}}C), \\ &= R\underline{\underline{H}}(\underline{\underline{H}}^{\#}[\sigma_{H}\underline{\underline{H}}])^{k}\underline{\underline{H}}^{\#}\underline{\underline{H}}C; \end{aligned}$$

by repeated application of (8.9),

$$\begin{aligned} &= R\underline{\underline{H}}(\underline{\underline{H}}^{\#}\underline{\underline{H}}\underline{\underline{Z}})^{k}\underline{\underline{H}}^{\#}\underline{\underline{H}}C \\ &= R\underline{\underline{S}}\underline{\underline{H}}(\underline{\underline{H}}^{\#}\underline{\underline{H}}\underline{\underline{Z}})^{k-1}H^{\#}HC, \\ &\quad\cdot \\ &\quad\cdot \\ &\quad\cdot \\ &= R\underline{\underline{S}}^{k}\underline{\underline{H}}\underline{\underline{H}}^{\#}\underline{\underline{H}}C, \\ &= R\underline{\underline{S}}^{k}\underline{\underline{H}}C, \\ &= R[\sigma_{H}^{k}\underline{\underline{H}}]C. \end{aligned}$$

The last equation calls for picking out the first p rows and the first m columns of $\sigma_{H}^{k}\underline{\underline{H}}$, which is just A_{1+k}, as required. □

(8.20) COMMENT. This is a considerably sharper result than Theorem (8.12), in two respects:

(i) It is no longer necessary to compute $\underline{\underline{Z}}$: we simply use the matrix $\underline{\underline{H}}_{\lambda',\lambda''}(\sigma_{A}\underline{\underline{A}})$, which is part of the data of the problem.

(ii) Formulas (8.18) give the desired realization in minimal form: there is no need to reduce (8.13) to a minimal realization (recall here (7.11)).

Notice also that the proof of (8.19) does not require (8.12) but depends (just like the latter) on direct use of (8.8).

An apparently serious limitation of the algorithm (8.16) is the necessity to verify *abstractly* that "$\underline{\underline{A}}$ has finite length". Of course, this can be done only on the basis of certain special hypotheses on $\underline{\underline{A}}$, given in advance. (Examples: (i) $A_k = 0$ for all $k > q$; (ii) A_k = coefficients of the Taylor expansion of a rational function.) Fortunately, the difficulty is only apparent, for the preceding developments can be sharpened further:

(8.21) FUNDAMENTAL THEOREM OF LINEAR REALIZATION THEORY. *Consider any infinite sequence* $\underline{\underline{A}}$ *and the corresponding Hankel matrix* $\underline{\underline{H}}$. *Suppose there exist integers* ℓ', ℓ'' *such that*

$$(8.22) \qquad \operatorname{rank} \underline{\underline{H}}_{\ell',\ell''}(\underline{\underline{A}}) = \operatorname{rank} \underline{\underline{H}}_{\ell'+1,\ell''}(\underline{\underline{A}}), \\ = \operatorname{rank} \underline{\underline{H}}_{\ell',\ell''+1}(\underline{\underline{A}}).$$

Then there exists unique extension $\hat{\underline{\underline{A}}}$ *of* $\underline{\underline{A}}$ *of order* $\ell' + \ell''$ *such that* $\lambda'_{\hat{A}} \leq \ell'$ *and* $\lambda''_{\hat{A}} \leq \ell''$; *moreover, applying formulas* (8.17-18) *with* $\lambda' = \ell'$, $\lambda'' = \ell''$ *gives a canonical realization of* $\hat{\underline{\underline{A}}}$.

PROOF. Exactly as in the necessity part of the proof of (8.8), condition (8.22) implies the existence of $\underline{\underline{S}}$ and $\underline{\underline{Z}}$ such that

$$(8.23) \qquad \sigma_{\underline{\underline{H}}}\underline{\underline{H}}_{\ell',\ell''}(\underline{\underline{A}}) = \underline{\underline{S}}\,\underline{\underline{H}}_{\ell',\ell''}(\underline{\underline{A}}) = \underline{\underline{H}}_{\ell',\ell''}(\underline{\underline{A}})\underline{\underline{Z}}.$$

Define an extension $\hat{\underline{\underline{A}}}$ of $\underline{\underline{A}}$ of order $\ell' + \ell''$ by

$$\sigma^k_{\underline{\underline{H}}}\underline{\underline{H}}_{\ell',\ell''}(\hat{\underline{\underline{A}}}) \triangleq \underline{\underline{S}}^k\underline{\underline{H}}_{\ell',\ell''}(\underline{\underline{A}}), \qquad k > 1.$$

By repeated application of (8.23), it follows that we have also

$$\sigma^{k}_{\underline{\underline{H}}}\underline{\underline{H}}_{\ell',\ell''}(\hat{\underline{\underline{A}}}) = \underline{\underline{H}}_{\ell',\ell''}(\underline{\underline{A}})\underline{\underline{Z}}^{k}, \qquad k \geq 0.$$

Now it is clear, from (8.8), that $\lambda'_{\hat{\underline{\underline{A}}}} \leq \ell'$ and $\lambda''_{\hat{\underline{\underline{A}}}} \leq \ell''$. The uniqueness of the extension follows immediately from (8.15). Moreover, Theorem (8.19) is still valid, even though (ℓ', ℓ'') is not necessarily minimal, because the proof of (8.19) depended only on (8.9) and not on the minimality of (ℓ', ℓ''). □

Theorem (8.21) says, in effect, that a canonical realization of some <u>extension</u> of $\underline{\underline{A}}$ is always possible as scon as (8.22) is satisfied. Moreover, (8.22) can be used as a <u>practical</u> criterion for constructing by trial and error a canonical realization of any $\underline{\underline{A}}$ known to have finite length (but without being given λ', λ'').

(8.24) EXAMPLES. (i) There is no scalar infinite sequence $(p = m = 1)$ $\underline{\underline{A}}$ for which (8.22) is <u>never</u> satisfied.

(ii) If $\underline{\underline{H}}_{\ell',\ell''}$ is square and has full rank (for instance, in the scalar case), then (8.22) is automatically satisfied.

(iii) If the algorithm (8.16) is applied <u>without</u> any information concerning condition (8.22), the system Σ defined by (8.18) will always realize <u>some</u> extension of $\underline{\underline{A}}$, at least of order 1. It is not known, however, how to get a simple formula which would determine the maximal order of this extension of $\underline{\underline{A}}$.

The remaining interesting question is then: What can be said if (8.22) is not satisfied for a <u>finite</u> amount of data $A_1, \ldots, A_N$ and any ℓ', ℓ'' satisfying $\ell' + \ell'' = N$. This problem is the topic of the next section.

(8.25) FINAL COMMENT. An essential feature of B. L. Ho's algorithm is that is preserves the block structure of the data of the problem. Of course, one can obtain parallel results by treating $\underline{\underline{H}}_{\ell', \ell''}$ as an ordinary matrix, disregarding its block-Hankel structure. Such a procedure requires looking at a minor of $\underline{\underline{H}}$ of maximum rank, and was described explicitly by SILVERMAN [1966] and SILVERMAN and MEADOWS [1969]. There does not seem to be any obvious computational advantage associated with the second method.

9. THEORY OF PARTIAL REALIZATIONS

In one obvious respect the theory of realizations developed in the previous section is rather unsatisfactory: it is concerned with infinite sequences. From here on we call a system satisfying (8.3) a *complete realization,* to distinguish it from the practically more interesting case given by

(9.1) DEFINITION. *Let* $\underline{\underline{A}} = (A_1, A_2, \dots)$ *be an infinite sequence of* $p \times m$ *matrices over a fixed field* K. *A dynamical system* $\Sigma = (F, G, H)$ *is a partial realization of order* r *of* $\underline{\underline{A}}$ *iff*

$$A_{k+1} = HF^kG \quad \underline{\text{for}} \quad k = 0, 1, \dots, r.$$

We shall use the same terminology if, instead of an infinite sequence $\underline{\underline{A}}$, we are given merely a finite sequence $\underline{\underline{A}}_s = (A_1, \dots, A_s)$, $s \geqq r$. The reason for this convention will be clear from the discussion to follow. We shall call the first r terms of $\underline{\underline{A}}$ a *partial sequence* (*of order* r).

The concepts of *canonical partial realization* and *minimal partial realization* will be understood in exactly the same sense as for a complete realization. We warn the reader, however, that now these two notions will turn out to be inequivalent, in that

minimal partial $\Rightarrow$ canonical partial

but not conversely.

Our main interest will be to determine all equivalence classes of minimal partial realizations; in general, a given sequence will have infinitely many inequivalent minimal partial realizations if r is sufficiently small.

According to the Main Theorem (8.21) of the theory of realizations, the minimal partial realization problem has a unique solution whenever the rank condition (8.22) is satisfied. If the length r of the partial sequence is prescribed a priori, it may well happen that (8.22) does not hold. What to do? Clearly, if we have a minimal partial realization (F, G, H) of order r we can extend the partial sequence of $\underline{\underline{A}}_r$ on which this realization is based to an infinite sequence canonically realized by (F, G, H) simply by setting

$$A_k \triangleq HF^{k-1}G, \quad k > r.$$

Consequently, we have the preliminary

(9.2) PROPOSITION. *The determination of a minimal partial realization for* $\underline{\underline{A}}_r$ *is equivalent to the determination of all extensions of a partial sequence* $\underline{\underline{A}}_r$ *such that the extended sequence is*

(i) *finite-dimensional and, more strongly,*

(ii) *its dimension is minimal in the class of all extensions.*

It is trivial to prove that finite-dimensional extensions exist for any partial sequence (of finite length). Hence the problem is immediately reduced to determining extensions which have minimal dimension. The solution of this latter problem consists of two steps. First, we show by a trivial argument that the minimal dimension can be bounded from below by an examination of the Hankel array defined by the partial sequence. Second, and this is rather surprising, we show that the lower bound can be actually attained. For further details, especially the characterization of equivalence classes of the minimal partial realizations, see KALMAN [1969c and 1970b].

(9.3) DEFINITION. *By the Hankel array* $\underline{\underline{H}}(\underline{\underline{A}}_r)$ *of a partial sequence* $\underline{\underline{A}}_r$ *we mean that* $r \times r$ *block Hankel matrix whose* $(i, j)^{th}$ *block is* A_{i+j-1} *if* $i + j - 1 \leq r$ *and undefined otherwise.*

In other words, the Hankel array of a partial sequence $\underline{\underline{A}}_r$ consists of block rows and columns made up of subsequences $A_p, \ldots, A_r$ $(1 \leq p \leq r)$ of $\underline{\underline{A}}_r$ and blank spaces.

(9.4) PROPOSITION. *Let* $n_o(\underline{\underline{A}}_r)$ *be the number of rows of the Hankel array of* $\underline{\underline{A}}_r$ *which are linearly independent of the rows above them. Then the dimension of a realization of* $\underline{\underline{A}}_r$ *is at least* $n_o(\underline{\underline{A}}_r)$.

PROOF. The rank of *any* Hankel matrix of an infinite sequence $\underline{\underline{A}}$ is a lower bound on the dimension of *any* realization of $\underline{\underline{A}}$, by Proposition (8.4). By Proposition (9.2), it suffices to consider a suitable extension $\underline{\underline{A}}$ of $\underline{\underline{A}}_r$. This implies "filling in" the blank spaces in the Hankel array of $\underline{\underline{A}}_r$. Regardless of how $\underline{\underline{H}}(\underline{\underline{A}}_r)$ is filled in, the rank of the resulting $r \times r$ block Hankel matrix is bounded from below by $n_o(\underline{\underline{A}}_r)$. □

By the block symmetry of the Hankel matrix, we would expect to be able to determine $n_o(\underline{\underline{A}}_r)$ by an analogous examination of the columns of the Hankel array of $\mathbf{A}_r$, thereby obtaining the *same* lower bound. This is indeed true. We prefer not to give a direct proof, since the result will follow as a corollary of the Main Theorem (9.7).

The critical fact is given by the

(9.5) MAIN LEMMA. *For a partial sequence* $\mathbf{A}_r$ *define*:

$\lambda'(\mathbf{A}_r)$ = *smallest integer such that for* $k' > \lambda'$ *every row of* $\mathbf{H}(\mathbf{A}_r)$ *is linearly dependent on the rows above it.*

$\lambda''(\mathbf{A}_r)$ = *smallest integer such that for* $k'' > \lambda''$ *every column in the* k-th *block column of* $\mathbf{H}(\mathbf{A}_r)$ *is linearly dependent on the columns to the left of it.*

Every partial sequence $\mathbf{A}_r$ *may be extended to an infinite sequence* $\mathbf{A}$ *in at least one way such that the condition*

(9.6) $$\text{rank } \mathbf{H}_{\mu,\nu}(\mathbf{A}) = n_o(\mathbf{A}_r) \quad \textit{for all} \quad \mu > \lambda'(\mathbf{A}_r), \ \nu > \lambda''(\mathbf{A}_r)$$

is satisfied.

PROOF. The existence of the numbers λ'. λ'' is trivial.

It suffices to show, for arbitrary r, how to select A_{r+1} in such a way that the numbers λ', λ'', and n_o remain constant.

Consider the first row of A_{r+1} and examine in turn all the first rows of the first, second, third, ..., λ'-th block rows in $\mathbf{H}(\mathbf{A}_r)$. If the first row of the first block row is linearly dependent on the rows above it (that is, 0), we fill in the first row of A_{r+1} using this linear dependence (that is, we make the first row of A_{r+1} all zeros). This choice of the first row of A_{r+1} will preserve linear dependencies for the first row of every block row below the second block row, by the definition of the Hankel pattern. If the first row in the first block row is linearly independent of those above (that is, contributes 1 to $n_o(\underline{\underline{A}}_r)$), we pass to the second block row and repeat the procedure. Eventually the first row of some block row will become linearly dependent on those above it, except when $\lambda' = r$; in that case, choose the first row of A_{r+1} to be linearly dependent of the first rows of $A_1, \ldots, A_r$. Repeating this process for the second, third, ... rows of each block row*, eventually A_{r+1} is determined without increasing λ' or n_o.

To complete the proof, we must show that the above definition of A_{r+1} also preserves the value of λ'. That is, we must show that no new independent columns are produced in the Hankel array of $\underline{\underline{A}}_r$ when A_{r+1} is filled in. This is verified immediately by noting that the definition of A_{r+1} implies the conditions

$$\operatorname{rank} \underline{\underline{H}}_{r,1} = \operatorname{rank} \underline{\underline{H}}_{r+1,1},$$

$$\operatorname{rank} \underline{\underline{H}}_{r-1,2} = \operatorname{rank} \underline{\underline{H}}_{r,2},$$

$$\cdots$$

$$\operatorname{rank} \underline{\underline{H}}_{1,r} = \operatorname{rank} \underline{\underline{H}}_{2,r} = \operatorname{rank} \underline{\underline{H}}_{1,r+1}. \qquad \square$$

---

*Of course, now linear dependence in the first step does not imply that the corresponding row of A_{r+1} will be all zeros.

With the aid of this simple but subtle observation, the problem is reduced to that covered by the Main Theorem (8.21) of Section 8. We have:

(9.7) MAIN THEOREM FOR MINIMAL PARTIAL REALIZATIONS.* *Let* $\underline{\underline{A}}_r$ *be a partial sequence. Then:*

(i) *Every minimal realization of* $\underline{\underline{A}}_r$ *has dimension* $n_o(\underline{\underline{A}}_r)$.

(ii) *All minimal realizations may be determined with the aid of B. L. Ho's formulas* (8.17-18) *with* $\lambda' = \lambda'(\underline{\underline{A}}_r)$ and $\lambda'' = \lambda''(\underline{\underline{A}}_r)$ *as given by Lemma* (9.5).

(iii) *If* $r \geqq \lambda'(\underline{\underline{A}}_r) + \lambda''(\underline{\underline{A}}_r)$ *then the minimal realization is unique. Otherwise there are as many minimal realizations as there are extensions of* $\underline{\underline{A}}_r$ *satisfying* (9.6).

PROOF. By the Main Lemma (9.5), every partial sequence $\underline{\underline{A}}_r$ has at least one infinite extension which preserves λ', λ'' and n_o. So we can apply the (8.21) of the preceding section.
It follows that the minimal partial realization is unique if $r \geqq \lambda'(\underline{\underline{A}}_r) + \lambda''(\underline{\underline{A}}_r)$ (the $\lambda'(\underline{\underline{A}}_r) + \lambda''(\underline{\underline{A}}_r) + 1$ Hankel matrix can be filled in completely with the available data); in the contrary case, the minimal extensions will depend on the manner in which the matrices $A_{r+1}, \dots, A_{\lambda'+\lambda''}$ have been determined (subject to the requirement (9.6)). □

In view of the theorem, we are justified in calling the integer $n_o(\underline{\underline{A}}_r)$ the *dimension* of $\underline{\underline{A}}_r$.

----

*A similar result was obtained simultaneously and independently by T. Tether (Stanford dissertation, 1969).

(9.8) REMARK. The essential point is that the quantities n_o, λ', and λ'' are uniquely determined already from *partial* data, irrespective of the possible nonuniqueness of the minimal extensions of the partial sequence. We warn, however, that this result does not generalize to all invariants of the minimal realization. For instance, one cannot determine from $\underline{\underline{A}}_r$ how many cyclic pieces a minimal realization of $\underline{\underline{A}}_r$ will have: some minimal realizations may be cyclic and others may not [KALMAN 1970b].

Finally, let us note also a second consequence of the Main Theorem:

(9.9) COROLLARY. *Suppose* $n_1(\underline{\underline{A}}_r)$ *is the number of independent columns of the Hankel array of* $\underline{\underline{A}}_r$ *(defined analogously with* $n_o(\underline{\underline{A}}_r)$*).* *Then* $\dim \underline{\underline{A}}_r = n_1(\underline{\underline{A}}_r)$.

PROOF. If $n_1(\underline{\underline{A}}_r) > n_o(\underline{\underline{A}}_r)$ then, using the Main Theorem, we get a contradiction to the fact that the rank of any Hankel matrix of an infinite sequence is lower bound for the dimension of any realization (Proposition (8.4)). If $n_1(\underline{\underline{A}}_r) < n_o(\underline{\underline{A}}_r)$ then extending $\underline{\underline{A}}_r$ to any $\underline{\underline{A}}_{\lambda'+\lambda''-1}$ we contradict the fact that rank $\underline{\underline{H}}_{\lambda',\lambda''}$ is at least equal to $n_o(\underline{\underline{A}}_r)$. □

In other words, the characteristic property of rank, that counting rank by row or column dependence yields identical results, is preserved even for incomplete Hankel arrays.

It is useful to check a simple case which illustrates some of the technicalities of the proof of the Main Lemma.

(9.10) EXAMPLE. The dimension of $(0, 0, \ldots, 0, A_r)$ is precisely $r \times \rho$, where $\rho = \text{rank } A_r$ and $\lambda' = \lambda'' = r$.

10. GENERAL THEORY OF OBSERVABILITY

In this concluding section, we wish to discuss the problem of observability in a rather general setting: we will not assume linearity, at least in the beginning. This is an ambitious program and leads to many more problems than results. Still, I think it is interesting to give some indication of the difficulties which are conceptual as well as mathematical. This discussion can also serve as an introduction to very recent research [KALMAN 1969a, 1970a] on the observability problem in certain classes of nonlinear systems.

The motivation for this section, as indeed for the whole theory of observability, stems from the writer's discovery [KALMAN 1960a] that the problem of (linear) statistical prediction and filtering can be formulated and resolved very effectively by consistent use of dynamical concepts and methods, and that this whole theory is a strict dual of the theory of optimal control of linear systems with quadratic Lagrangian. For those who are familiar with the standard classical theory of statistical filtering (see, for instance, YAGLOM [1962]), we can summarize the situation very simply by saying that

Wiener-Kolmogorov filter
+ theory of finite-dimensional linear dynamical systems
= Kalman filter.

For the latter, the original papers are [KALMAN 1960a, 1963a] and [KALMAN and BUCY 1961].

The reader interested in further details and a modern exposition is referred especially to the monograph of KALMAN [1969b].

We shall examine here only one aspect of this theory (which does not involve any stochastic elements): the strict formulation of the "duality principle" between reachability and observability. This principle was formally stated for the first time by KALMAN [1960c], but the pertinent discussion in this paper is limited to the linear case and is somewhat ad-hoc. Aided by research progress since 1960, it is now possible to develop a completely general approach to the "duality principle". We shall do this and, as a by-product, we shall obtain a new and strictly deductive proof of the principle in the now classical linear case.

We shall introduce a general notion of the "dual" system, and use it to replace the problem of observability by an equivalent problem of reachability. In keeping with the point of view of the earlier lectures, we shall view a system in terms of its input/output map f and dualize f (rather than Σ). The constructibility problem will not be of direct interest, since its theory is similar to that of the observability problem.

Let Ω, Γ be the same sets as defined in Section 4 and used from then on. We assume that both Ω and Γ are K-vector spaces (K = arbitrary field) and recall the definition of the shift operators σ_Ω and σ_Γ on Ω and Γ (see (3.10)). We denote both shift operators by z but ignore, until later, the $K[z]$-module structure on Ω and Γ.

By a constant (not necessarily linear) input/output map $f: \Omega \to \Gamma$ we shall mean any map f which commutes with the shift operators, that is,

$$f(z\cdot\omega) = z\cdot(f(\omega)).$$

Let us now formulate the general problem of this section:

(10.1) PROBLEM OF OBSERVABILITY. Given an input/output map f, its canonical realization Σ, and an input sequence $\nu \in \Omega$ applied after $t = 0$. Determine the state x of Σ at $t = 0$ from the knowledge of the output sequence of Σ after $t = 0$.

This problem cannot be solved in general! To see this, recall that the state set X_f of f may be viewed as a set of functions

$$\{f(\omega\circ\cdot)(1): \Omega \to K^p: \nu \mapsto f(\omega\circ\nu)(1)\}$$

since ω' is Nerode-equivalent to ω iff

$$f(\omega'\circ\cdot)(1) = f(\omega\circ\cdot)(1)$$

Giving $\nu \in \Omega$ and the corresponding output sequence amounts to giving various values of $f(\omega\circ\cdot)(1)$ (namely those corresponding to the sequences $\emptyset$, ν_r, $z\nu_r + \nu_{r-1}$, ..., ν, $z\nu$, $z^2\nu$, ...), and it may happen that these substitutions do not yield enough values of the function $f(\omega\circ\cdot)(1)$ to determine the function itself. This situation has been recognized for a long time in automata theory, where, in an almost self-explanatory terminology, one says that "Σ is initial-state determinable by an infinite multiple experiment (possibly infinitely many different ν's) but not necessarily by a single experiment (single ν chosen at will)." See MOORE [1956]. The problem is further complicated by the fact that it may make a difference whether or not we have a free choice of ν. KALMAN, FALB, and ARBIB [1969, Section 6.3)] give some related comments.

A further difficulty inherent in the preceding discussion is that the problem is posed on a purely set-theoretic level and does not lend itself to the introduction of more refined structural assumptions. We shall therefore reformulate the problem in such a way as to *focus attention on determining those properties of the initial state which can be computed from the combined knowledge of the input and output sequence occurring after* $t = 0$.

For simplicity, we shall fix the value of ν at 0 (no loss of generality, since f is not linear). Then the output sequence resulting from x after $t = 0$ is given simply as $f(\omega)$, where $x = [\omega]_f$.

We shall use the circumflex to denote certain classes of functions from a set into the field K. *For the moment, this class will be the class of all functions*. Thus

$$\hat{\Gamma} = \{\text{all functions } \Gamma \to K\}.$$

An element $\hat{\gamma}$ of $\hat{\Gamma}$ is simply a "rule" (in practice, a computing algorithm) which assigns to each possible output sequence γ in Γ

a number in the field K. If γ resulted from the state $x = [\omega]_f$, then

$$\hat{\gamma}(\gamma) = \hat{\gamma}(f(\omega)) = (\hat{\gamma} \circ f)(\omega)$$

gives the value of a certain function in $\hat{\Omega}$ and, by definition of the state, also the value of a certain function in $\hat{X}$. This suggests the

(10.2) DEFINITION. An element $\hat{x} \in \hat{X}$ is an observable costate iff there is a $\hat{\gamma}_{\hat{x}} \in \hat{\Gamma}$ such that we have identically for all $\omega \in \Omega$

$$\hat{x}([\omega]_f) = \hat{\gamma}_{\hat{x}}(f(\omega)).$$

In other words, no matter what the initial state $x = [\omega]_f$ is, the value of $\hat{x}$ at x can always be determined by applying the rule $\hat{\gamma}_{\hat{x}}$ to the output sequence $f(\omega)$ resulting from x. Note, carefully, that this definition subsumes (i) a fixed choice of the class of functions denoted by the circumflex, and (ii) a fixed input sequence after $t = 0$ (here $\nu = 0$). For certain purposes, it may be necessary to generalize the definition in various ways [KALMAN 1970 a], but here we wish to avoid all unessential complications.

According to Definition (10.2), we shall see that a system is completely observable iff every costate is observable. This agrees with the point of view adopted earlier (see Section 4) in an ad-hoc fashion. Also, the vague requirement to "determine x" used in

(10.1) is now replaced by a precise notion which can be manipulated (via the actual definition of the circumflex) to express limitations on the algorithms that we may apply to the output sequence of the system.

The requirement "every costate is observable" can be often replaced by a much simpler one. For instance, if X is a vector space, it is enough to know that "every linear costate is observable" or even just that "every element of some dual basis is an observable costate"; if X is an algebraic variety, it is natural to interpret "complete observability" as "every element of the coordinate ring of X is an observable costate" [KALMAN 1970a].

We can now carry out a straightforward "dualization" of the setup involved in the definition of the input/output map $f\colon \Omega \to \Gamma$. First, we adopt (again with respect to a fixed interpretation of the circumflex):

(10.3) DEFINITION. The dual of an input/output map $f\colon \Omega \to \Gamma$ is the map

$$\hat{f}\colon \hat{\Gamma} \to \hat{\Omega}\colon \hat{\gamma} \mapsto \hat{\gamma} \circ f$$

Note that $\hat{f}$ is well-defined, since the circumflex means the class of all functions.

As to the next step, we wish to prove that constancy is inherited under dualization. To do this, we have to induce a definition of the shift operator on $\hat{\Gamma}$ and $\hat{\Omega}$. The only possible definitions are the obvious ones:

$$\sigma_{\hat{\Gamma}}\colon\ \hat{\Gamma} \to \hat{\Gamma}\colon\ \hat{\gamma} \mapsto [\sigma_{\hat{\Gamma}}\hat{\gamma}\colon\ \gamma \mapsto \hat{\gamma}(\sigma_{\Gamma}\gamma)];$$

$$\sigma_{\hat{\Omega}}\colon\ \hat{\Omega} \to \hat{\Omega}\colon\ \hat{\omega} \mapsto [\sigma_{\hat{\Omega}}\hat{\omega}\colon\ \omega \mapsto \hat{\omega}(\sigma_{\Omega}\omega)].$$

Both of these new shift operators will be denoted by z^{-1}. The reason for this notation will become clear later.

Now it is easy to verify:

(10.4) PROPOSITION. *If* f *is constant, so is* $\hat{f}$.

PROOF. We apply the definitions in suitable sequence:

$$\begin{aligned}
\hat{f}(z^{-1}\cdot\hat{\gamma})(\omega) &= (z^{-1}\cdot\hat{\gamma})(f(\omega)) && (\text{def. of } \hat{f}),\\
&= \hat{\gamma}(z\cdot f(\omega)) && (\text{def. of } \sigma_{\hat{\Gamma}}),\\
&= \hat{\gamma}(f(z\cdot\omega)) && (f \text{ is constant}),\\
&= \hat{f}(\hat{\gamma})(z\cdot\omega) && (\text{def. of } \hat{f}),\\
&= (z^{-1}\cdot\hat{f}(\hat{\gamma}))(\omega) && (\text{def. of } \sigma_{\hat{\Omega}}),
\end{aligned}$$

and so we see that $\hat{f}$ commutes with z whenever f does. □

At this stage, we cannot as yet view $\hat{f}$ as the input/output map of a dynamical system because concatenation is not yet defined on $\hat{\Gamma}$, and therefore $\hat{\Gamma}$ is not yet a properly defined "input set". In other words, it is necessary to check that the notion of time is also inherited under dualization. In general, this does not appear to be possible without some strong limitation on the class $\hat{\Gamma}$. Here we shall look only at the simplest

(10.5) HYPOTHESIS. *Every function* $\hat{\gamma}$ *in* $\hat{\Gamma}$ *satisfies the finiteness condition:* *There is an integer* $|\hat{\gamma}|$ (*dependent on* $\hat{\gamma}$) *such that for all* $\gamma, \delta \in \Gamma$ *the condition*

$$\gamma_k = \delta_k, \; k = 1, \ldots, |\hat{\gamma}|$$

implies

$$\hat{\gamma}(\gamma) = \hat{\gamma}(\delta).$$

In other words, we assume that the value of each $\hat{\gamma}$ at γ is uniquely determined by some *finite portion* of the output sequence γ.

Assuming (10.5), it is immediate that $\hat{\Gamma}$ admits a concatenation multiplication which corresponds (at least intuitively) to the usual one defined on Ω:

$$(10.6) \qquad \hat{\gamma}\circ\hat{\delta} = z^{-|\hat{\delta}|}\cdot\hat{\gamma} + \hat{\delta}.$$

We can now prove the expected theorem, which may be regarded as the precise form of the "duality" principle:

(10.7) THEOREM. *Let* f *be an arbitrary constant input/output map and* $\hat{f}$ *its dual*. *Suppose further that* (10.5) *holds*. *Then each observable costate of* f (*relative to* $\hat{\Gamma}$ *satisfying* (10.5)) *may be viewed as a reachable state of* $\hat{f}$, *and conversely*.

PROOF. First we determine the Nerode equivalence classes on $\hat{\Gamma}$ induced by $\hat{f}$. By definition

$$\hat{\delta} \in (\hat{\gamma})_{\hat{f}} \quad \text{iff} \quad \hat{f}(\hat{\delta}\circ\hat{\epsilon}) = \hat{f}(\hat{\gamma}\circ\hat{\epsilon})$$

for all $\hat{\epsilon} \in \hat{\Gamma}$. Now $\hat{f}$ is linear (!); in fact, direct use of the definition of $\hat{f}$ and (10.6) gives

$$\hat{\delta} \in (\hat{\gamma})_{\hat{f}} \quad \text{iff} \quad (\hat{\gamma}\circ f)(\omega) \;=\; (\hat{\delta}\circ f)(\omega),\ \omega \in \Omega.$$

So $\hat{\gamma}\circ f$ and $\hat{\delta}\circ f$ are equal as elements of $\check{X}$: they define the same observable costate. In fancier language, the assignment

$$(10.8) \qquad d\colon\ X_{\hat{f}} \to \hat{X}_f\colon\ (\hat{\gamma})_{\hat{f}} \mapsto \hat{\gamma}\circ f$$

is well defined and constitutes a bijection between the reachable states of $\hat{f}$ and those costates of f which are observable relative to the function class ^. □

Thus (10.5) is a sufficient condition for the duality principle to hold. However, the fact that the canonical realization of $\hat{f}$ is completely reachable is not quite the same as saying that the canonical realization of f is completely observable because the latter <u>depends</u> on the choice of $\hat{\Gamma}$ and therefore is not an intrinsic property of f. Moreover, Theorem (10.7) does not give any indication how "big" $X_{\hat{f}}$ is and it may certainly happen that the observability problem for f is much more difficult than the reachability problem. These matters will be illustrated later by some examples.

Now we deduce the original form of the duality principle from Theorem (10.7). The essential point is that (10.5) holds automatically as a result of linearity.

New definition of the function class: let the circumflex denote the class of all <u>K-linear</u> functions. (All the underlying sets with the K-vector spaces, so the definition makes sense.)

The following facts are well known:

(10.9) PROPOSITION. *Let* $*$ *denote duality in the sense of* K-*vector spaces*. *Then*:

$$\hat{\Gamma} \triangleq (K^p[[z^{-1}]])^* = K^p[z^{-1}],$$

$$\hat{\Omega} \triangleq (K^m[z])^* = K^m[[z]].$$

Now we can state the

(10.10) MAIN THEOREM. *Suppose* f *is* K-*linear, constant, finite-dimensional*. *Suppose further that* ^ *means* K-*linear duality*. *Then*:

(i) $\hat{f}$ *is* K-*linear and constant, that is, a* $K[z^{-1}]$-*homomorphism* (*and therefore written as* f^*) *and finite-dimensional*.

(ii) *The reachable states of* f^* *are isomorphic with the* K-*linear dual of* X_f; *hence every costate of* X_f *is observable*.

PROOF. The fact that Γ is K-linear implies, by (10.3), that $\hat{f}$ is K-linear; the constancy of f always implies that of $\hat{f}$, by Proposition (10.4). (*Caution*: $\hat{f}$ is *not* the K[z]-linear dual of the K[z]-homomorphism f, and the construction given here cannot be simplified. See Remark (4.26A).)

To prove the second part, we note that by Proposition (10.9) Hypothesis (10.5) holds and thus $\hat{f} = f^*$ is a well-defined input/output map of a dynamical system. We must prove that the reachable states of f^* are isomorphic with X_f^*, the K-linear dual of X_f. This amounts to proving that the K-vector space of functions

$$x \mapsto \hat{\gamma}(h_f(x),\ h_f(z\cdot x),\ \dots\ )$$

is isomorphic with the K-vector space X_f^*. It suffices to prove that the K-vector space generated by the K-linear functions

$$(10.11) \quad \{\lambda:\ x \mapsto [h_f(z^i \cdot x)]_j, \quad i = 0, 1, \dots \text{ and } j = 1, \dots, m\}$$

is isomorphic with X_f^*. Suppose that, for fixed x, every $\lambda(x) = 0$. Then $x = 0$, by definition of the Nerode equivalence relation induced by f (recall here the discussion from Section 3). Since X_f is finite-dimensional by hypothesis, it follows from this property of the functions $\{\lambda\}$ that they generate X_f^*. Obviously, $\dim X_f^* = \dim X_f$, so that everything is proved. □

In other terms, the fact that $f = K[z]$-homomorphism together with the appropriate definition of ^ implies that

$$\hat{f}:\ K^p[z^{-1}] \to K^m[[z]]$$

is a $K[z^{-1}]$-homomorphism. Since (10.5) holds, we can interpret $\hat{f}$ in a system-theoretic way, as follows: the output of the dual system at $t = -k$ due to input $\hat{\gamma}$ is given by the assignment

$$\hat{\gamma} \mapsto \hat{f}(\hat{\gamma})(-k),$$

which is a linear function defined on the k-th term of the input sequence. In fact, we have

$$\begin{aligned}\hat{\gamma}(\gamma) &= \hat{f}(\hat{\gamma})(\omega),\\ &= (\hat{\gamma}\circ f)(\omega),\\ &= \sum_k (\hat{f}(\hat{\gamma})(-k))(\omega_k).\end{aligned}$$

(10.12) REMARK. It is essentially a consequence of Proposition (10.9) that $\hat{f}$ turns out to be the same kind of algebraic object as f. Note, however, that

> <u>under duality the input and output terminals are interchanged and</u> t <u>is replaced by</u> -t (<u>hence</u> z <u>by</u> z^{-1}).

In terms of the pictorial definition of a system, this statement simply amounts to "reversing the directions of the arrows", which is the "right" way to define duality in the most general mathematical context, namely in category theory. We would expect that the duality principles of system theory will eventually become a part of this very general duality theory. This has not happened yet because the correct categories to be considered in the study of dynamical systems have not yet been determined. It is likely that eventually many different categories will have to be looked at in studying dynamical problems.

We shall now present an example which should help to interpret the previous results. We emphasize, however, that the theory sketched here is still in a very rudimentary form.

(10.13) EXAMPLE. Consider the system Σ defined by

$$x(t+1) = 2x(t) + u(t), \; y(t) = x(t), \; t \in \underline{\underline{Z}};$$

$$y(t) = \begin{cases} 0 & \text{if } 0 \leqq x(t) < 1/2, \\ 1 & \text{if } 1/2 \leqq x(t) < 1, \end{cases}$$

with $X = U = Y = \underline{\underline{R}} \bmod 1$, i.e., the interval $[0, 1)$. (1 is to be thought of as identified with 0.) We let $u(t) = 0$. We view x through its binary representation

$$x = \sum_{k=0}^{\infty} \xi_k(x) 2^{-k}, \qquad \xi_k(x) = 0 \text{ or } 1.$$

It is clear from the definition of the system that the output sequence due to any x is precisely

$$\gamma_x = (\xi_1(x), \xi_2(x), \ldots).$$

If x is irrational, infinitely many terms are needed to identify it. Consequently, the x's are isomorphic with the Nerode equivalence classes induced by f_Σ. So Σ cannot be reduced.

Relative to "^ = functions", every costate of f_Σ is observable, provided that Hypothesis (10.5) is <u>not</u> satisfied. If it is, then only those costates defined on fixed-length rationals are observable (more precisely, these are functions which depend only on a fixed finite subset of the $\xi_k(x)$'s). Thus: <u>either $\hat{f}$ does not define a dynamical system or not all costates are observable</u>.

Now let us replace the set $[0, 1)$ by its intersection with the rationals. It is clear that there is now a <u>finite</u> algorithm for determining x: we simply apply the results of partial realization theory of the previous section. (We take $K = \underline{\underline{Z}}_2$ and the problem is to express x from $(\xi_1(x), \ldots, \xi_2(x)0$ as a ratio of polynomials in $\underline{\underline{Z}}_2[2]$--which is always possible since each x is rational.) However, x is not "effectively computable" in the strict sense since there is no way of knowing when the algorithm has stopped. In other words, given an arbitrary costate $\hat{x}$ there exists no fixed rule $\hat{\gamma}_{\hat{x}}$ such that the application of $\hat{\gamma}_{\hat{x}}$ to γ_x gives $\hat{x}(x)$ for all x. On the other hand, substituting into $\hat{x}$ the results of the partial-realization algorithm will give an approximation to the value of $\hat{x}(x)$ which always converges in a finite (but a priori unknown) number of steps as more values of the output sequence are observed. In short, the costate-determination algorithm has certain pseudo-random elements in it and therefore cannot be described through the machinery of deterministic dynamical systems. (Is there some relation here to the conceptual difficulties of Quantum Mechanics?)

11. HISTORICAL COMMENTS

It is not an exaggeration to say that the entire theory of linear, constant (and here, discrete-time) dynamical systems can be viewed as a systematic development of the equivalent algebraic conditions (2.8) and (2.15).

Of course, the use of modules (over $K[z]$) to study a constant square matrix (see (4.13)) has been "standard" since the 1920's under the influence of E. NOETHER and especially after the publication of the Modern Algebra of VAN DER WAERDEN. Condition (2.15), by itself, must be also quite old. For instance, GANTMAKHER [1959, Vol. 1, p. 203] attributes to KRYLOV [1931] the idea of computing the characteristic polynomial of a square matrix A by choosing a random vector b and computing successively $b, Ab, A^2b, \dots$ until linear dependence is obtained, which yields the coefficients of $\det (zI - A)$. (The method will succeed iff X_A is cyclic with generator g.) However, the merger of (4.13) with (2.15), which is the essential idea in the algebraic theory of linear systems, was done explicitly first in KALMAN [1965b].

We shall direct our remarks here mainly to the history of conditions (2.8) and (2.15) as related to controllability. See also earlier comments in KALMAN [1960c, pp. 481, 483, 484] and in KALMAN, HO, and NARENDRA [1963, pp. 210-212]. We will have to bear in mind that the development of modern control theory cannot be separated from the development of the concept of controllability; moreover, the technological problems of the 1950's and even earlier had a major influence on the genesis of mathematical ideas (just as the latter have led to many new technological applications of control in the 1960's).

R.E.Kalman

The writer developed the mathematical definition of controllability with applications to control theory, during the first part of 1959. (Unpublished course notes at Johns Hopkins University, 1958/59.) These first definitions were in the form of (2.17) and (2.3). Formal presentations of the results were made in Mexico City (September, 1959, see KALMAN [1960b]), University of California at Berkeley (April, 1969, see KALMAN [1960d]), and Moskva (June, 1960, see KALMAN [1960c]), and in scientific lectures on many other concurrent occasions in the U.S. As far as the writer is aware, a conscious and explicit definition of controllability which combines a control-theoretic wording with a precise mathematical criterion was first given in the above references. There are of course many instances of similar ideas arising in related contexts. Perhaps the comments below can be used as the starting point of a more detailed examination of the situation in a seminar in the history of ideas.

The following is the chain of the writer's own ideas culminating in the publications mentioned above:

(1) In KALMAN [1954] it is pointed out (using transform methods) that continuous-time linear systems can be controlled by a linear discrete-time (sampled-data) controller in *finite time*.*

*It is sometimes claimed in the mathematical literature of optimal control theory that this *cannot* be done with a linear system. This is false; the correct statement is "cannot be done with a linear controller producing control functions which are *continuous (and not merely piecewise continuous!) in time*." Such a restriction is completely irrelevant from the technological point of view. As a matter of fact, computer-controlled systems have been proposed and built for many years on the basis of *linear, time-optimal* control.

(2) Transposing the result of KALMAN [1954] from transfer functions to state variables, an algorithm was sketched for the solution of the discrete-time time-optimal control of systems with bounded control and linear continuous-time dynamics. [KALMAN, 1957]

(3) As a popularization of the results of the preceding work, the same technique was applied to give a general method for the design of linear sampled-data systems by KALMAN and BERTRAM [1958].

Some background comments concerning these papers are appropriate:

(1) The ideas and method presented in KALMAN [1954] descend directly from earlier (and very well known) engineering research on time-optimal control. (The main references in KALMAN [1954] are: McDONALD [1950], HOPKIN [1951], BOGNER and KAZDA [1954], as well as a research report included in KALMAN [1955].) Although the results of KALMAN [1954] on <u>linear</u> time-optimal control were considered to be new when published, it became clear later that similar ideas were at least implicit in OLDENBOURG and SARTORIUS [1951, §90, p. 219] and in TSYPKIN's work in the early 1950's. The engineering idea of nonlinear time-optimal control goes back, at least, to DOLL [1943] and to OLDENBURGER in 1944, although the latter's work was unfortunately not widely known before 1957. During the same time, there was much interest in the same problems in other countries; see, for instance, FELDBAUM [1953] and UTTLEY and HAMMOND [1953]. Mathematical work in these problems probably began with BUSHAW's dissertation [1952] in which, to quote from KALMAN [1955, before equation (40)], "...[it was] rigorously proved that the intuition which led to the formulation of the [engineering] theory [quoted above] was indeed correct." TSIEN's survey [1954] contains a lengthy account of this state

of affairs and was ready by many. We emphasize: none of this extensive literature contains even a hint of the algebraic considerations related to controllability.

(2-3) The critical insight gained and recorded in KALMAN [1957] is the following: the solution of the discrete-time time-optimal control problem is equivalent to expressing the state as a linear combination of a certain vector sequence (related to control and dynamics) with coefficients bounded by 1 in absolute value, the coefficients being the values of the optimal control sequence. The linear independence of the first n vectors of the sequence guarantees that every point in a neighborhood of zero can be moved to the origin in at most n steps (hence the terminology of "complete controllability"); and the condition for this is identical with (2.17) (stated in KALMAN [1957] and KALMAN and BERTRAM [1958] only for the case $\det F \neq 0$ and $m = 1$). A thorough discussion of these matters is found in KALMAN [1960c; see especially Theorem I, p. 485]. A serious conceptual error in KALMAN [1957] occurred, however, in that complete controllability was not assumed, as a hypothesis for the existence of time-optimal control law, but an attempt was made to show that the controllability is almost always complete [Lemma 1]. In fact, this lemma is true, with a small technical modification in the condition. Only much later did it become clear (see the discussion of Theorem D in the Introduction), however, that a dynamical system is always completely controllable (in the nonconstant case, completely reachable) if it is derived from an external description. It was this difficulty, very mysterious in 1957, which led to the development of a formal machinery for the definition of controllability during the next two years. The changing point of view is already apparent in KALMAN and BERTRAM [1958]; the unpublished paper promised there was delayed precisely because the algebraic machinery to prove Theorem D was out of reach in 1957-8. Consult also the findings of the bibliographer RUDOLF [1969].

IN SUMMARY: *under the stimulation of the engineering problems of minimal-time optimal control*, the researches begun by KALMAN [1954, 1957] and KALMAN and BERTRAM [1958] eventually evolved into what has come to be called the *mathematical theory of controllability* (of linear systems).

Beginning about 1955, and *stimulated by the same engineering problems*, PONTRYAGIN and his school in the USSR developed their *mathematical theory of optimal control* around the celebrated "Maximum Principle". (They were well aware of the survey of TSIEN [1954] mentioned above, and referenced it both in English and in the Russian translation of 1956.) We now know that *any* theory of control, regardless of its particular mathematical style, must contain ingredients related to controllability. So it is interesting to examine how explicitly the controllability condition appears in the work of PONTRYAGIN and related research.

GAMKRELIDZE [1957, §2; 1958 §1, §2] calls the time optimal control problem associated with the system

$$(11.1) \qquad dx/dt = Ax + bu(t)$$

"nondegenerate" iff b is not contained in a proper A-invariant subspace of R^n. He notes immediately that this is equivalent to

$$(11.2) \qquad \det (b, Ab, \ldots, A^{n-1}b) \neq 0$$

(i.e., the special case of (2.8) for $m = 1$). He then proves: in the "degenerate" case the problem either reduces to a simpler one or the motion cannot be influenced by the control function $u(\cdot)$. All this is very close to an explicit definition of controllability. However, in discussing the general case $m > 1$, GAMKRELIDZE [1958, §3, Section 1] defines "nondegeneracy" of the system

$$(11.3) \qquad dx/dt = Ax + Bu(t)$$

as the condition

$$(11.4) \qquad \det (b_i, Ab_i, \ldots, A^{n-1}b_i) \neq 0 \quad \text{for every column } b_i \in B,$$

but he does not show that this generalized condition of "nondegeneracy" for (11.3) inherits the interesting characterization proved for "nondegeneracy" in the case of (11.1). In fact, condition (11.4) is much too strong to prove this; the correct condition is (2.8), that is, complete controllability. In other words, in GAMKRELIDZE's work (11.4) plays the role of a technical condition for eliminating "degeneracy" (actually, lack of uniqueness) from a particular optimal control problem and is not explicitly related to the more basic notion of complete controllability. Neither GAMKRELIDZE nor PONTRYAGIN [1958] give an interpretation of (11.4) as a property of the dynamical system (11.3), but employ (11.4) only in relation to the particular problem of time-optimal control. See also KALMAN [1960c, p. 484]. A similar point of view is taken by LaSALLE [1960]; he calls a dynamical system (11.3) satisfying (2.8) "proper" but then goes on to require (11.4) (to assure the uniqueness of the time-optimal controls) and calls such systems "normal".

The assumption of some kind of "nondegeneracy" condition was apparently unavoidable in the early phases of research on the time-optimal control problem. For example, ROSE [1953, pp. 39-58] examines this problem for (11.1); by defining "nondegeneracy" [p. 41] by a condition equivalent ot (11.2), he obtains most of GAMKRELIDZE's results in the special case when A has real eigenvalues [Theorem 12]. ROSE uses determinants closely related to the now familiar lemmas in controllability theory but he, too, fails to formulate controllability as a concept independent of the time-optimal control problem.

A similar situation exists in the calculus of variations. The so-called Caratheodory classes (after CARATHEODORY [1933]) correspond to a kind of classification of controllability properties of nonconstant systems. In fact, the standard notion of a normal family of extremals of the calculus of variations is closely related to condition (11.4), suitably generalized via (2.5) to nonconstant systems.* Normality is used in the calculus of variations mainly as a "nondegeneracy" condition.

It is important to note that the "nondegeneracy" conditions employed in optimal control and the calculus of variations play mainly the role of eliminating annoying technicalities and simplifying proofs.

----

*The use of the word "normal" by LaSALLE [1960] for (11.4) is only accidentally coincident with the earlier use of the "normal" in the calculus of variations.

With suitable formulation, however, the basic results of time-optimal control theory continue to hold without the assumption of complete controllability. The same is not true, however, of the four kinds of theorems mentioned in the Intorduction, and therefore these results are more relevant to the story of controllability than the time-optimal control discussed above.

There is a considerable body of literature relevant to controllability theory which is quite independent of control theory. For instance, the treatment of a reachability condition in partial differential equations goes back at least to CHOW [1940] but perhaps it is fairer to attribute it to Caratheodory's well-known approach to entropy via the nonintegrability condition. The current status of these ideas as related to controllability is reviewed by WEISS [1969, Section 9]. An independent and very explicit study of reachability is due to ROXIN [1960]; unfortunately, his examples were purely geometric and therefore the paper did not help in clarifying the celebrated condition (2.8). The Wronskian determinant of the classical theory of ordinary differential equations with variable coefficients also has intersections with controllability theory, as pointed out recently with considerable success by SILVERMAN [1966]. Many problems in control theory were misunderstood or even incorrectly solved before the advent of controllability theory. Some of these are mentioned in KALMAN [1963b, Section 9]. For relations with automata theory, see ARBIB [1965].

Let us conclude by stating the writer's own current position as to the significance of controllability as a subject in mathematics:

(1) Controllability is basically an algebraic concept. (This claim applies of course also to the nonlinear controllability results obtained via the Pfaffian method.)

(2) The historical development of controllability was heavily influenced by the interest prevailing in the 1950's in optimal control theory. Ultimately, however, controllability is seen as a relatively minor component of that theory.

(3) Controllability as a conceptual tool is indispensable in the discussion of the relationship between transfer functions and differential equations and in questiohs relating to the four theorems of the Introduction.

(4) The chief current problem in controllability theory is the extension to more elaborate algebraic structures.

For a survey of the historical background of observability, which would take us too far afield here, the reader should consult KALMAN [1969b].

12. REFERENCES

Section A: General References

M. A. ARBIB

[1965] A common framework for automata theory and control theory, SIAM J. Contr., 3:206-222.

C. W. CURTIS and I. REINER

[1962] Representation Theory of Finite Groups and Associative Algebras, Interscience-Wiley.

E. M. DAY and A. D. WALLACE

[1967] Multiplication induced in the state space of an act, Math. System Theory, 1:305-314.

C. A. DESOER and P. VARAIYA

[1967] The minimal realization of a nonanticipative impulse response matrix, SIAM J. Appl. Math., 15:754-764.

E. G. GILBERT

[1963] Controllability and observability in multivariable control systems, SIAM J. Control, 1:128-151.

B. L. HO and R. E. KALMAN

[1966] Effective construction of linear state-variable models from input/output functions, Regelungstechnik, 14:545-548.

[1969] The realization of linear, constant input/output maps, I. Complete realizations, SIAM J. Contr., to appear.

S. T. HU

[1965] Elements of Modern Algebra, Holden-Day.

R. E. KALMAN

[1960a] A new approach to linear filtering and prediction problems, J. Basic Engr. (Trans. ASME), 82D:35-45.

[1960b] Contributions to the theory of optimal control, Bol. Soc. Mat. Mexicana, 5:102-119.

[1960c] On the general theory of control systems, Proc. 1st IFAC Congress, Moscow; Butterworths, London.

[1962] Canonical structure of linear dynamical systems, Proc. Nat. Acad. of Sci. (USA), 48:596-600.

[1963a] New methods in Wiener filtering theory, Proc. 1st Symp. on Engineering Applications of Random Function Theory and Probability, Purdue University, November 1960, pp 270-388, Wiley. (Abridged from RIAS Technical Report 61-1.)

[1963b] Mathematical description of linear dynamical systems, SIAM J. Contr., 1:152-192.

[1965a] Irreducible realizations and the degree of a rational matrix, SIAM J. Contr., 13:520-544.

[1965b] Algebraic structure of linear dynamical systems. I. The Module of Σ, Proc. Nat. Acad. Sci. (USA), 54:1503-1508.

[1967] Algebraic aspects of the theory of dynamical systems, in Differential Equations and Dynamical Systems, J. K. Hale and J. P. LaSalle (eds.), pp. 133-146, Academic Press.

[1969a] On multilinear machines, J. Comp. and System Sci., to appear.

[1969b] Dynamic Prediction and Filtering Theory, Springer, to appear.

[1969c] On partial realizations of a linear input/output map, Guillemin Anniversary Volume, Holt, Winston and Rinehart.

[1970a] Observability in multilinear systems, to appear.

[1970b] The realization of linear, constant, input/output maps. II. Partial realizations, SIAM J. Control, to appear.

R. E. KALMAN and R. S. BUCY

[1961] New results in linear prediction and filtering theory, J. Basic Engr. (Trans. ASME, Ser. D), 83D:95-100.

R. E. KALMAN, P. L. FALB and M. A. ARBIB

[1969] Topics in Mathematical System Theory, McGraw-Hill.

R. E. KALMAN, Y. C. HO and K. NARENDRA

[1963] Controllability of linear dynamical systems, Contr. to Diff. Equations, 1:189-213.

C. E. LANGENHOP

[1964] On the stabilization of linear systems, Proc. Am. Math. Soc., 15:735-742.

[1965] Algebra, Addison-Wesley.

S. MAC LANE

[1963] Homology, Springer.

L. A. MARKUS

[1965] Controllability of nonlinear processes, SIAM J. Control, 3:78-90.

E. F. MOORE

[1956] Gedanken-experiments on sequential machines, in Automata Studies, C. E. Shannon and J. McCarthy (eds.), pp. 129-153, Princeton University Press.

P. MUTH

[1899] Theorie und Anwendung der Elementartheiler, Teubner, Leipzig.

A. NERODE

[1958] Linear automaton transformations, Proc. Amer. Math. Soc., 9:541-544.

L. SILVERMAN

[1966] Representation and realization of time-variable linear systems, Doctoral dissertation, Columbia University.

L. M. SILVERMAN and H. E. MEADOWS

[1969] Equivalent realizations of linear systems, SIAM J. Control, to appear.

H. WEBER

[1898] Lehrbuch der Algebra, Vol. 1, 2nd Edition, reprinted by Chelsea, New York.

L. WEISS

[1969] Lectures on Controllability and Observability, C.I.M.E. Seminar.

L. WEISS and R. E. KALMAN

[1965] Contributions to linear system theory, Intern. J. Engr. Sci., 3:141-171.

W. M. WONHAM

[1967] On pole assignment in multi-input controllable linear systems, IEEE Trans. Auto. Contr., AC-12:600-665.

A. M. YAGLOM

[1962] An Introduction to the Theory of Stationary Random Functions, Prentice-Hall.

D. C. YOULA

[1966] The synthesis of linear dynamical systems from prescribed weighting patterns, SIAM J. Appl. Math., 14:527-549.

D. C. YOULA and P. TISSI

[1966] n-port synthesis via reactance extraction, Part I, IEEE Intern. Convention Record.

O. ZARISKI and P. SAMUEL

[1958] Commutative Algebra, Vol. 1, Van Nostrand.

Section B: References for Section 11

M. A. ARBIB

[1965] A common framework for automata theory and control theory, SIAM. J. Contr., 3:206-222.

I. BOGNER and L. F. KAZDA

[1954] An investigation of the switching criteria for higher order contactor servomechanisms, Trans. AIEE, 73 II:118-127.

D. W. BUSHAW

[1952] Differential equations with a discontinuous forcing term, doctoral dissertation, Princeton University.

C. CARATHEODORY

[1933] Über die Einteilung der Variationsprobleme von Lagrange nach Klassen, Comm. Mat. Helv., 5:1-19.

W. L. CHOW

[1940] Über Systeme von linearen partiellen Differentialgleichungen erster Ordnung, Math. Annalen, :98-105.

H. G. DOLL

[1943] Automatic control system for vehicles, US Patent 2,463,362.

A. A. FELDBAUM

[1953] Avtomatika i Telemekhanika, 14:712-728.

R. V. GAMKRELIDZE

[1957] On the theory of optimal processes in linear systems (in Russian), Dokl. Akad. Nauk SSSR, 116:9-11.

[1958] The theory of optimal processes in linear systems (in Russian), Izvestia Akad. Nauk SSSR, 2:449-474.

F. R. GANTMAKHER

[1959] The Theory of Matrices, 2 vols., Chelsea.

A. M. HOPKIN

[1951] A phase-plane approach to the compensation of saturating servomechanisms, Trans. AIEE, 70:631-639.

R. E. KALMAN

[1954] Discussion of a paper by Bergen and Ragazzini, Trans. AIEE, 73 II: 245-246.

[1955] Analysis and design principles of second and higher-order saturating servomechanisms, Trans. AIEE, 74 II:294-310.

[1957] Optimal nonlinear control of saturating systems by intermittent control, IRE WESCON Convention Record, 1 IV:130-135.

[1960b] Contributions to the theory of optimal control, Bol. Soc. Mat. Mexicana, 5:102-119.

[1960c] On the general theory of control systems, Proc. 1st IFAC Congress, Moscow; Butterworths, London.

[1960d] Lecture notes on control system theory (by M. Athans and G. Lendaris), Univ. of Calif. at Berkeley.

[1963b] Mathematical description of linear dynamical systems, SIAM J. Contr., 1:152-192.

[1965b] Algebraic structure of linear dynamical systems. I. The Module of Σ, Proc. Nat. Acad. Sci. (USA), 54:1503-1508.

[1969b] Dynamic Prediction and Filtering Theory, Springer, to appear.

R. E. KALMAN and J. E BERTRAM

[1958] General synthesis procedure for computer control of single and multi-loop linear systems, Trans, AIEE, 77 III:602-609.

R. E. KALMAN, Y. C. HO and K. NARENDRA

[1963] Controllability of linear dynamical systems, Contr. to Diff. Equations, 1:189-213.

A. N. KRYLOV

[1931] On the numerical solution of the equation by which the frequency of small oscillations is determined in technical problems (in Russian), Izv. Akad. Nauk SSSR Ser. Fix.-Mat., 4:491-539.

J. P. LaSALLE

[1960] The time-optimal control problem, Contr. Nonlinear Oscillations, Vol. 5, Princeton Univ. Press.

D. C. McDONALD

[1950] Nonlinear techniques for improving servo performance, Proc. Nat. Electronics Conf. (USA), 6:400-421.

R. C. OLDENBOURG and H. SARTORIUS

[1951] Dynamik selbstättiger Regelungen, 2nd edition, Oldenbourg, Munchen.

R. OLDENBURGER

[1957] Optimum nonlinear control, Trans. ASME, 79:527-546.

[1966] Optimal and Self-Optimizing Control, MIT Press.

L. S. PONTRYAGIN

[1958] Optimal control processes (in Russian), Uspekhi Mat. Nauk, 14:3-20.

N. J. ROSE

[1953] Theoretical aspects of limit control, Report 459, Stevens Institute of Tech., Hoboken, N.J.

E. ROXIN

[1960] Reachable zones in autonomous differential systems, Bol. Soc. Mat. Mexicana, 5:125-135.

K. E. RUDOLF

[1969] On some unpublished works of R. E. Kalman, not to be unpublished.

H. S. TSIEN

[1954] Engineering Cybernetics, McGraw-Hill.

A. M. UTTLEY and P. H. HAMMOND

[1953] The stabilization of on-off controlled servomechanisms, in Automatic and Manual Control, Academic Press.

L. WEISS

[1969] Lectures on Controllability and Observability, C.I.M.E. Seminar